



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 0 786 519 A2**

(12)

**EUROPEAN PATENT APPLICATION**

(43) Date of publication:  
**30.07.1997 Bulletin 1997/31**

(51) Int Cl.<sup>6</sup>: **C12N 15/00**

(21) Application number: **97100117.7**

(22) Date of filing: **07.01.1997**

(84) Designated Contracting States:  
**AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC  
NL PT SE**  
Designated Extension States:  
**AL LT LV RO SI**

(30) Priority: **05.01.1996 US 9861**

(71) Applicant: **HUMAN GENOME SCIENCES, INC.**  
**Rockville, MD 20850-3338 (US)**

(72) Inventors:  
• **Kunsch, Charles A.**  
**Gaithersburg, Maryland 20882 (US)**

- **Choi, Gil H.**  
**Rockville, Maryland 20850 (US)**
- **Barash, Steven C.**  
**Rockville, Maryland 20850 (US)**
- **Dillon, Patrick J.**  
**Gaithersburg, Maryland 20879 (US)**
- **Fannon, Michael R.**  
**Silver Spring, Maryland 20906 (US)**
- **Rosen, Craig A.**  
**Laytonsville, Maryland 20882 (US)**

(74) Representative: **VOSSIUS & PARTNER**  
**Postfach 86 07 67**  
**81634 München (DE)**

**(54) Staphylococcus aureus polynucleotides and sequences**

(57) The present invention provides polynucleotide sequences of the genome of *Staphylococcus aureus*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynu-

cleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.

**EP 0 786 519 A2**



## Description

The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Staphylococcus aureus*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation, polypeptide production, assays and pharmaceutical development, among others.

The genus *Staphylococcus* includes at least 20 distinct species. (For a review see Novick, R. P., *The Staphylococcus as a Molecular Genetic System*, Chapter 1, pgs. 1-37 in *MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI*, R. Novick, Ed., VCH Publishers, New York (1990)). Species differ from one another by 80% or more, by hybridization kinetics, whereas strains within a species are at least 90% identical by the same measure.

The species *Staphylococcus aureus*, a gram-positive, facultatively aerobic, clump-forming cocci, is among the most important etiological agents of bacterial infection in humans, as discussed briefly below.

## Human Health and *S. Aureus*

*Staphylococcus aureus* is a ubiquitous pathogen. (See, for instance, Mims *et al.*, *MEDICAL MICROBIOLOGY*, Mosby-Year Book Europe Limited, London, UK (1993)). It is an etiological agent of a variety of conditions, ranging in severity from mild to fatal. A few of the more common conditions caused by *S. aureus* infection are burns, cellulitis, eyelid infections, food poisoning, joint infections, neonatal conjunctivitis, osteomyelitis, skin infections, surgical wound infection, scalded skin syndrome and toxic shock syndrome, some of which are described further below.

### Burns

Burn wounds generally are sterile initially. However, they generally compromise physical and immune barriers to infection, cause loss of fluid and electrolytes and result in local or general physiological dysfunction. After cooling, contact with viable bacteria results in mixed colonization at the injury site. Infection may be restricted to the non-viable debris on the burn surface ("eschar"), it may progress into full skin infection and invade viable tissue below the eschar and it may reach below the skin, enter the lymphatic and blood circulation and develop into septicaemia. *S. aureus* is among the most important pathogens typically found in burn wound infections. It can destroy granulation tissue and produce severe septicaemia.

### Cellulitis

Cellulitis, an acute infection of the skin that expands from a typically superficial origin to spread below the cutaneous layer, most commonly is caused by *S. aureus* in conjunction with *S. pyogenes*. Cellulitis can lead to systemic infection. In fact, cellulitis can be one aspect of synergistic bacterial gangrene. This condition typically is caused by a mixture of *S. aureus* and microaerophilic streptococci. It causes necrosis and treatment is limited to excision of the necrotic tissue. The condition often is fatal.

### Eyelid infections

*S. aureus* is the cause of styes and of sticky eye\* in neonates, among other eye infections. Typically such infections are limited to the surface of the eye, and may occasionally penetrate the surface with more severe consequences.

### Food poisoning

Some strains of *S. aureus* produce one or more of five serologically distinct, heat and acid stable enterotoxins that are not destroyed by digestive process of the stomach and small intestine (enterotoxins A-E). Ingestion of the toxin, in sufficient quantities, typically results in severe vomiting, but not diarrhoea. The effect does not require viable bacteria. Although the toxins are known, their mechanism of action is not understood.

### Joint infections

*S. aureus* infects bone joints causing diseases such as osteomyelitis.

### Osteomyelitis

*S. aureus* is the most common causative agent of haematogenous osteomyelitis. The disease tends to occur in

children and adolescents more than adults and it is associated with non-penetrating injuries to bones. Infection typically occurs in the long end of growing bone, hence its occurrence in physically immature populations. Most often, infection is localized in the vicinity of sprouting capillary loops adjacent to epiphyseal growth plates in the end of long, growing bones.

#### *Skin infections*

*S. aureus* is the most common pathogen of such minor skin infections as abscesses and boils. Such infections often are resolved by normal host response mechanisms, but they also can develop into severe internal infections. Recurrent infections of the nasal passages plague nasal carriers of *S. aureus*.

#### *Surgical Wound Infections*

Surgical wounds often penetrate far into the body. Infection of such wound thus poses a grave risk to the patient. *S. aureus* is the most important causative agent of infections in surgical wounds. *S. aureus* is unusually adept at invading surgical wounds; sutured wounds can be infected by far fewer *S. aureus* cells than are necessary to cause infection in normal skin. Invasion of surgical wound can lead to severe *S. aureus* septicaemia. Invasion of the blood stream by *S. aureus* can lead to seeding and infection of internal organs, particularly heart valves and bone, causing systemic diseases, such as endocarditis and osteomyelitis.

#### *Scalded Skin Syndrome*

*S. aureus* is responsible for "scalded skin syndrome" (also called toxic epidermal necrosis, Ritter's disease and Lyell's disease). This disease occurs in older children, typically in outbreaks caused by flowering of *S. aureus* strains produce exfoliation (also called scalded skin syndrome toxin). Although the bacteria initially may infect only a minor lesion, the toxin destroys intercellular connections, spreads epidermal layers and allows the infection to penetrate the outer layer of the skin, producing the desquamation that typifies the disease. Shedding of the outer layer of skin generally reveals normal skin below, but fluid lost in the process can produce severe injury in young children if it is not treated properly.

#### *Toxic Shock Syndrome*

Toxic shock syndrome is caused by strains of *S. aureus* that produce the so-called toxic shock syndrome toxin. The disease can be caused by *S. aureus* infection at any site, but it is too often erroneously viewed exclusively as a disease solely of women who use tampons. The disease involves toxemia and septicaemia, and can be fatal.

#### *Nocosomal Infections*

In the 1984 National Nosocomial Infection Surveillance Study ("NNIS") *S. aureus* was the most prevalent agent of surgical wound infections in many hospital services, including medicine, surgery, obstetrics, pediatrics and newborns.

#### *Resistance to drugs of S. aureus strains*

Prior to the introduction of penicillin the prognosis for patients seriously infected with *S. aureus* was unfavorable. Following the introduction of penicillin in the early 1940s even the worst *S. aureus* infections generally could be treated successfully. The emergence of penicillin-resistant strains of *S. aureus* did not take long, however. Most strains of *S. aureus* encountered in hospital infections today do not respond to penicillin; although, fortunately, this is not the case for *S. aureus* encountered in community infections.

It is well known now that penicillin-resistant strains of *S. aureus* produce a lactamase which converts penicillin to penicillinoic acid, and thereby destroys antibiotic activity. Furthermore, the lactamase gene often is propagated episomally, typically on a plasmid, and often is only one of several genes on an episomal element that, together, confer multidrug resistance.

Methicillins, introduced in the 1960s, largely overcame the problem of penicillin resistance in *S. aureus*. These compounds conserve the portions of penicillin responsible for antibiotic activity and modify or alter other portions that make penicillin a good substrate for inactivating lactamases. However, methicillin resistance has emerged in *S. aureus*, along with resistance to many other antibiotics effective against this organism, including aminoglycosides, tetracycline, chloramphenicol, macrolides and lincosamides. In fact, methicillin-resistant strains of *S. aureus* generally are multiply drug resistant.

The molecular genetics of most types of drug resistance in *S. aureus* has been elucidated (See Lyon *et al.*, *Microbiology Reviews* 51: 88-134 (1987)). Generally, resistance is mediated by plasmids, as noted above regarding penicillin resistance; however, several stable forms of drug resistance have been observed that apparently involve integration of a resistance element into the *S. aureus* genome itself.

Thus far each new antibiotic gives rise to resistance strains, strains emerge that are resistance to multiple drugs and increasingly persistent forms of resistance begin to emerge. Drug resistance of *S. aureus* infections already poses significant treatment difficulties, which are likely to get much worse unless new therapeutic agents are developed.

### Molecular Genetics of *Staphylococcus Aureus*

Despite its importance in, among other things, human disease, relatively little is known about the genome of this organism.

Most genetic studies of *S. aureus* have been carried out using the strain NCTC8325, which contains prophages psi11 psi12 and psi13, and the UV-cured derivative of this strain, 8325-4 (also referred to as RN450), which is free of the prophages.

These studies revealed that the *S. aureus* genome, like that of other *staphylococci*, consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory genetic elements, such as prophages, plasmids, transposons and the like.

Physical characterization of the genome has not been carried out in any detail. Pattee *et al.* published a low resolution and incomplete genetic and physical map of the chromosome of *S. aureus* strain NCTC 8325. (Pattee *et al.* Genetic and Physical Mapping of Chromosome of *Staphylococcus aureus* NCTC 8325, Chapter 11, pgs. 163-169 in. MOLECULAR BIOLOGY OF THE STAPHYLOCOCCI, R.P. Novick, Ed., VCH Publishers, New York, (1990) The genetic map largely was produced by mapping insertions of Tn551 and Tn4001, which, respectively, confer erythromycin and gentamicin resistance, and by analysis of SmaI-digested DNA by Pulsed Field Gel Electrophoresis ("PFGE").

The map was of low resolution; even estimating the physical size of the genome was difficult, according to the investigators. The size of the largest SmaI chromosome fragment, for instance, was too large for accurate sizing by PFGE. To estimate its size, additional restriction sites had to be introduced into the chromosome using a transposon containing a SmaI recognition sequence.

In sum, most physical characteristics and almost all of the genes of *Staphylococcus aureus* are unknown. Among the few genes that have been identified, most have not been physically mapped or characterized in detail. Only a very few genes of this organism have been sequenced. (See, for instance Thornsberry, J., *Antimicrobial Chemotherapy* 21 Suppl C: 9-16 (1988), current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. aureus* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. aureus* infection involves the programmed expression of *S. aureus* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. aureus* genes and genomic organization would dramatically improve understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. aureus* would provide reagents for, among other things, detecting, characterizing and controlling *S. aureus* infections. There is a need therefore to characterize the genome of *S. aureus* and for polynucleotides and sequences of this organism.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS: 1-5,191.

The present invention provides the nucleotide sequence of several thousand contigs of the *Staphylococcus aureus* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-5,191.

The present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

The nucleotide sequence of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence which is at least 95%, preferably 99% and most preferably 99.9%, identical to the nucleotide sequence of SEQ ID NOS:1-5,191 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

The present invention further provides systems, particularly computer-based systems which contain the sequence

information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Staphylococcus aureus* genome.

Another embodiment of the present invention is directed to fragments, preferably isolated fragments, of the *Staphylococcus aureus* genome having particular structural or functional attributes. Such fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs," fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs," and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter referred to as diagnostic fragments or "DFs."

Each of the ORFs in fragments of the *Staphylococcus aureus* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Staphylococcus aureus* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to polypeptides and proteins, preferably isolated polypeptides and proteins, encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention can from cells which have been altered to express them.

The invention further provides polypeptides, preferably isolated polypeptides, comprising *Staphylococcus aureus* epitopes and vaccine compositions comprising such polypeptides. Also provided are methods for vaccinating an individual against *Staphylococcus aureus* infection.

The invention further provides methods of obtaining homologs of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs or antigens of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, antigens, or one of the DFs of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies, antigens or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b) determining whether the agent binds to said protein.

The present genomic sequences of *Staphylococcus aureus* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Staphylococcus aureus* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Staphylococcus aureus* researchers and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes

has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

**FIGURE 1** is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

**FIGURE 2** is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Staphylococcus aureus* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Staphylococcus aureus* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq\_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research (TIGR) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading frames (ORFs) is accomplished by processing contigs with zorf. The ORFs are searched against *S. aureus* sequences from Genbank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990)). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

The present invention is based on the sequencing of fragments of the *Staphylococcus aureus* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-5,191. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

In addition to the aforementioned *Staphylococcus aureus* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-5,191, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-5,191" refers to any portion of the SEQ ID NOS:1-5,191 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Staphylococcus aureus* open reading frames (ORFs), expression modulating fragment (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in sample ("DFs"). A non-limiting identification of preferred representative fragments is provided in Tables 1-3.

As discussed in detail below, the information provided in SEQ ID NOS:1-5,191 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Staphylococcus aureus* proteins.

While the presently disclosed sequences of SEQ ID NOS:1-5,191 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-5,191. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-5,191 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-5,191 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-5,191 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191.

As discussed elsewhere hererin, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining

libraries and for sequencing are provided below, for instance. A wide variety of *Staphylococcus aureus* strains that can be used to prepare *S. aureus* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC®).

The nucleotide sequences of the genomes from different strains of *Staphylococcus aureus* differ somewhat. However, the nucleotide sequences of the genomes of all *Staphylococcus aureus* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-5,191. Nearly all will be at least 99% identical and the great majority will be 99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-5,191 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444 (1988) can be used to generate the percent identity of nucleotide sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

### COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-5,191 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, "provided" refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; *i.e.*, a nucleotide sequence provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-5,191. Such a manufacture provides a large portion of the *Staphylococcus aureus* genome and parts thereof (*e.g.*, a *Staphylococcus aureus* open reading frame (ORF)) in a form which allows a skilled artisan to examine the manufacture using means not directly applicable to examining the *Staphylococcus aureus* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-5,191 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Staphylococcus aureus* genome which contain homology to ORFs or proteins from both *Staphylococcus aureus* and from other organisms. Among the ORFs discussed

herein are protein encoding fragments of the *Staphylococcus aureus* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Staphylococcus aureus* genome.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Staphylococcus aureus* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the *Staphylococcus aureus* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990), was used to identify open reading frames within the *Staphylococcus aureus* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for



accessing and processing the genomic sequence (such as search tools, comparing tools, etc.) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

## 5 BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to fragments of the *Staphylococcus aureus* genome, preferably to isolated fragments. The fragments of the *Staphylococcus aureus* genome of the present invention include, but are not limited to fragments which encode peptides, hereinafter open reading frames (ORFs), fragments which  
10 modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Staphylococcus aureus* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Staphylococcus aureus* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification  
15 means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS: 1-5,191, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These  
20 include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Staphylococcus aureus* DNA can be mechanically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate an *Staphylococcus aureus* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS: 1-5,191. Well  
25 known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library of *Staphylococcus aureus* genomic DNA. Thus, given the availability of SEQ ID NOS: 1-5,191, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS: 1-5,191 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

Tables 1, 2 and 3 list ORFs in the *Staphylococcus aureus* genomic contigs of the present invention that were  
35 identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive or more selective lists.

Table 1 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are at least 80 amino acids long and over a continuous region of at least 50 bases which are 95% or more identical (by BLAST analysis) to an *S. aureus* nucleotide sequence available through Genbank in November 1996.

Table 2 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through Genbank by September 1996.

Table 3 sets out ORFs in the *Staphylococcus aureus* contigs of the present invention that do not match significantly,  
45 by BLASTP analysis, a polypeptide sequence available through Genbank by September 1996.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column indicates the reading frame, taking the first 5' nucleotide of the contig as the start of the +1 frame; the fourth column indicates the first nucleotide of the ORF, counting from the 5' end of the contig strand; and the fifth column indicates the length of each ORF in nucleotides.

In Tables 1 and 2, column six, lists the Reference\* for the closest matching sequence available through Genbank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column seven in Tables 1 and 2 provides the gene name\* of the matching sequence; column eight provides the BLAST identity\* score from the comparison of the ORF and the homologous gene; and column nine  
55 indicates the length in nucleotides of the highest scoring segment pair\* identified by the BLAST identity analysis.

In Table 3, the last column, column six, indicates the length of each ORF in amino acid residues.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (e.g., at positions



1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.*, possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list per cent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the per cent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Staphylococcus aureus* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Staphylococcus aureus* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Staphylococcus aureus* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below.

A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Staphylococcus aureus* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Staphylococcus aureus* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-5,191, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably 99% and most preferably 99.9% identical to SEQ ID NOS:1-5,191, with a sequence from another isolate of the same species.

Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the nucleic acid sequences mentioned above. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated.

Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.*, sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Staphylococcus aureus* origin isolated by using part or all of the fragments in question as a probe or primer.

Each of the ORFs of the *Staphylococcus aureus* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particular *Staphylococcus aureus*. Especially preferred in this regard are ORF such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Staphylococcus aureus*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Sta-*

*phylococcus aureus*, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* **6**: 3073 (1979); Cooney *et al.*, *Science* **241**: 456 (1988); and Dervan *et al.*, *Science* **251**: 1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* **56**: 560 (1991) and OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988).

The present invention further provides recombinant constructs comprising one or more fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Staphylococcus aureus* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK and KS (+ and -), pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG (available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

The present invention further provides host cells containing any one of the isolated fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation, which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Staphylococcus aureus* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (e.g., an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immu-

no-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (e.g., microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, e.g., *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Staphylococcus aureus* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

"Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancers and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL, 2<sup>nd</sup> Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *Staphylococcus aureus*, *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*. Others

may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23: 175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

An additional aspect of the invention includes *Staphylococcus aureus* polypeptides which are useful as immunodiagnostic antigens and/or immunoprotective vaccines, collectively "immunologically useful polypeptides". Such immunologically useful polypeptides may be selected from the ORFs disclosed herein based on techniques well known in the art and described elsewhere herein. The inventors have used the following criteria to select several immunologically useful polypeptides:

As is known in the art, an amino terminal type I signal sequence directs a nascent protein across the plasma and outer membranes to the exterior of the bacterial cell. Such outer membrane polypeptides are expected to be immunologically useful. According to Izard, J. W. et al., *Mol. Microbiol.* 13, 765-773; (1994), polypeptides containing type I signal sequences contain the following physical attributes: The length of the type I signal sequence is approximately 15 to 25 primarily hydrophobic amino acid residues with a net positive charge in the extreme amino terminus; the central region of the signal sequence must adopt an alpha-helical conformation in a hydrophobic environment; and the region surrounding the actual site of cleavage is ideally six residues long, with small side-chain amino acids in the -1 and -3 positions.

Also known in the art is the type IV signal sequence which is an example of the several types of functional signal sequences which exist in addition to the type I signal sequence detailed above. Although functionally related, the type IV signal sequence possesses a unique set of biochemical and physical attributes (Strom, M. S. and Lory, S., *J. Bacteriol.* 174, 7345-7351; 1992). These are typically six to eight amino acids with a net basic charge followed by an additional sixteen to thirty primarily hydrophobic residues. The cleavage site of a type IV signal sequence is typically after the initial six to eight amino acids at the extreme amino terminus. In addition, all type IV signal sequences contain a phenylalanine residue at the +1 site relative to the cleavage site.

Studies of the cleavage sites of twenty-six bacterial lipoprotein precursors has allowed the definition of a consensus amino acid sequence for lipoprotein cleavage. Nearly three-fourths of the bacterial lipoprotein precursors examined contained the sequence L-(A,S)-(G,A)-C at positions -3 to +1, relative to the point of cleavage (Hayashi, S. and Wu, H. C. *Lipoproteins in bacteria*. *J. Bioenerg. Biomembr.* 22, 451-471; 1990).

It well known that most anchored proteins found on the surface of gram-positive bacteria possess a highly conserved carboxy terminal sequence. More than fifty such proteins from organisms such as *S. pyogenes*, *S. mutans*, *E. faecalis*, *S. pneumoniae*, and others, have been identified based on their extracellular location and carboxy terminal amino acid sequence (Fischetti, V. A. *Gram-positive commensal bacteria deliver antigens to elicit mucosal and systemic immunity*. *ASM News* 62, 405410; 1996). The conserved region is comprised of six charged amino acids at the extreme carboxy terminus coupled to 15-20 hydrophobic amino acids presumed to function as a transmembrane domain. Immediately adjacent to the transmembrane domain is a six amino acid sequence conserved in nearly all proteins ex-

aminated. The amino acid sequence of this region is L-P-X-T-G-X, where X is any amino acid.

Amino acid sequence similarities to proteins of known function by BLAST enables the assignment of putative functions to novel amino acid sequences and allows for the selection of proteins thought to function outside the cell wall. Such proteins are well known in the art and include "lipoprotein", "periplasmic", or "antigen".

An algorithm for selecting antigenic and immunogenic *Staphylococcus aureus* polypeptides including the foregoing criteria was developed by the present inventors. Use of the algorithm by the inventors to select immunologically useful *Staphylococcus aureus* polypeptides resulted in the selection of several ORFs which are predicted to be outer-membrane-associated proteins. These proteins are identified in Table 4, below, and shown in the Sequence Listing as SEQ ID NOS:5,192 to 5,255. Thus the amino acid sequence of each of several antigenic *Staphylococcus aureus* polypeptides listed in Table 4 can be determined, for example, by locating the amino acid sequence of the ORF in the Sequence Listing. Likewise the polynucleotide sequence encoding each ORF can be found by locating the corresponding polynucleotide SEQ ID in Tables 1, 2, or 3, and finding the corresponding nucleotide sequence in the sequence listing.

As will be appreciated by those of ordinary skill in the art, although a polypeptide representing an entire ORF may be the closest approximation to a protein found *in vivo*, it is not always technically practical to express a complete ORF *in vitro*. It may be very challenging to express and purify a highly hydrophobic protein by common laboratory methods. As a result, the immunologically useful polypeptides described herein as SEQ ID NOS:5,192-5,255 may have been modified slightly to simplify the production of recombinant protein, and are the preferred embodiments. In general, nucleotide sequences which encode highly hydrophobic domains, such as those found at the amino terminal signal sequence, are excluded for enhanced *in vitro* expression of the polypeptides. Furthermore, any highly hydrophobic amino acid sequences occurring at the carboxy terminus are also excluded. Such truncated polypeptides include for example the mature forms of the polypeptides expected to exist in nature.

Those of ordinary skill in the art can identify soluble portions the polypeptide identified in Table 4, and in the case of truncated polypeptides sequences shown as SEQ ID NOS:5,192-5,255, may obtain the complete predicted amino acid sequence of each polypeptide by translating the corresponding polynucleotides sequences of the corresponding ORF listed in Tables 1,2 and 3 and found in the sequence listing.

Accordingly, polypeptides comprising the complete amino acid of an immunologically useful polypeptide selected from the group of polypeptides encoded by the ORFs identified in Table 4, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto, and most preferably at least 99% identical thereto form an embodiment of the invention; in addition polypeptides comprising an amino acid sequence selected from the group of amino acid sequences shown in the sequence listing as SEQ ID NOS:5,191-5,255, or an amino acid sequence at least 95% identical thereto, preferably at least 97% identical thereto and most preferably at least 99% identical thereto, form an embodiment of the invention. Polynucleotides encoding the foregoing polypeptides also form part of the present invention.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention, particularly those epitope-bearing portions (antigenic regions) identified in Table 4. The epitope-bearing portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G., Shinnick, T. M., Green, N. and Leamer, R. A. (1983) "Antibodies that react with predetermined sites on proteins", Science, 219:660-666. Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. See, for instance, Wilson et al., Cell 37:767-778 (1984) at 777.

Antigenic epitope-bearing peptides and polypeptides of the invention preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. Non-limiting examples of antigenic polypeptides or peptides that can be used to generate *S. aureus* specific antibodies include: a polypeptide comprising peptides shown in Table 4 below. These polypeptide fragments have been determined to bear antigenic epitopes of indicated *S. aureus* proteins by the analysis of the Jameson-Wolf antigenic index, a representative sample of which is shown in Figure 3.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means. See, e.g., Houghten, R. A. (1985) General method for the rapid solid-phase synthesis of large numbers of peptides:

specificity of antigen-antibody interaction at the level of individual amino acids. Proc. Natl. Acad. Sci. USA 82: 5131-5135; this "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten et al. (1986). Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. See, for instance, Geysen et al., supra. Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. et al. (1996) on Per-alkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Table 4 lists immunologically useful polypeptides identified by an algorithm which locates novel *Staphylococcus aureus* outermembrane proteins, as is described above. Also listed are epitopes or "antigenic regions" of each of the identified polypeptides. The antigenic regions, or epitopes, are delineated by two numbers x-y, where x is the number of the first amino acid in the open reading frame included within the epitope and y is the number of the last amino acid in the open reading frame included within the epitope. For example, the first epitope in ORF 168-6 is comprised of amino acids 36 to 45 of SEQ ID NO:5,192, as is described in Table 4. The inventors have identified several epitopes for each of the antigenic polypeptides identified in Table 4. Accordingly, forming part of the present invention are polypeptides comprising an amino acid sequence of one or more antigenic regions identified in Table 4. The invention further provides polynucleotides encoding such polypeptides.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Staphylococcus aureus*, of the fragments of the *Staphylococcus aureus* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Staphylococcus aureus* is defined as a homolog of a fragment of the *Staphylococcus aureus* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Staphylococcus aureus* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among these are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-5,191 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ ID NOS:1-5,191 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis et al., PCR PROTOCOLS, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-5,191 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, one skilled in the art will recognize that by employing high stringency conditions (e.g., annealing at 50-60°C in 6X SSPE and 50% formamide, and washing at 50- 65°C in 0.5X SSPE) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency

conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

When using DNA probes derived from SEQ ID NOS:1-5,191, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-5,191, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (e.g., hybridizing at 50-65°C in 5X SSPC and 50% formamide, and washing at 50-65°C in 0.5X SSPC), sequences having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (e.g., hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Staphylococcus aureus*.

## ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and industrial purposes consistent with the type of putative identification of the polypeptide. Such identifications permit one skilled in the art to use the *Staphylococcus aureus* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., Macmillan Publications, Ltd. NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESSES, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

### 1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Staphylococcus aureus* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-5,191.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*, *Symbiosis* 21: 79 (1986) and Voragen *et al.* in BIOCATALYSTS IN AGRICULTURAL BIOTECHNOLOGY, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series* 389: 93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Staphylococcus aureus*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1: 21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in GENE MANIPULATIONS AND FUNGI; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for



quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta.* 872: 83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose-produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, *Acid Proteinases Structure Function and Biology*, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, *Industrial Enzymes*, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, *Report Industrial Enzymes* by 1990, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London* 310:227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other hand, has been described in detail by Bud *et al.*, *Chemistry in Britain* (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo-selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination. A variety of commercially important enzymes have previously been isolated from members of *Staphylococcus aureus*. These include Sau3A and Sau96I.

## 2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *MONOCLONAL ANTIBODY TECHNOLOGY: LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256: 495-497 (1975)), the trioma technique, the human B- cell hybridoma technique (Kozbor *et al.*, *Immunology Today*



4: 72 (1983), pgs. 77-96 of Cole *et al.*, in MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc. (1985)).

Any animal (mouse, rabbit, *etc.*) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175: 109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above-described antibodies in detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labelling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym.* 62:308 (1979); Engval, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J. W. J. *Immunol. Meth.* 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays to identify cells or tissues in which a fragment of the *Staphylococcus aureus* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, *Meth. Enzym.* 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays as well as for immunoaffinity purification of the proteins of the present invention.

### 3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using one of the DFs, antigens or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies, or one or more of the DFs, or one or more antigens of the present invention and assaying for binding of the DFs, antigens or antibodies to components within the test sample.

Conditions for incubating a DF, antigen or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs, antigens or antibodies of the present invention. Examples of such assays can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry*; PCT publication WO95/32291, and *Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985), all of which are hereby incorporated herein by reference.

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based

on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the Dfs, antigens or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound DF, antigen or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody, antigen or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed Dfs, antigens and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

#### 4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Staphylococcus aureus* fragment and contigs herein described.

In general, such methods comprise steps of:

- (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Staphylococcus aureus* genome; and
- (b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, Nucl. Acids Res. 6:3073 (1979); Cooney *et al.*, Science 241:456 (1988); and Dervan *et al.*, Science 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated

to be effective in model systems. Information contained in the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

## 5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be used to modulate the growth or pathogenicity of *Staphylococcus aureus*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth or pathogenicity of *Staphylococcus aureus* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth or pathogenicity by binding to an important protein thus blocking the biological activity of the protein, while other agents may bind to a component of the outer surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of vaccines derived from membrane associated polypeptides are well known in the art. The inventors have identified particularly preferred immunogenic *Staphylococcus aureus* polypeptides for use as vaccines. Such immunogenic polypeptides are described above and summarized in Table 4, below.

As used herein, a "related organism" is a broad term which refers to any organism whose growth or pathogenicity can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can

be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s) serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in admixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, e. g., human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16<sup>th</sup> Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

## 6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present disclosure.

**ILLUSTRATIVE EXAMPLES****LIBRARIES AND SEQUENCING****1. Shotgun Sequencing Probability Analysis**

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2: 231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability,  $P_0$ , that any given base in a sequence of size  $L$ , in nucleotides, is not sequenced after a certain amount,  $n$ , in nucleotides, of random sequence has been determined can be calculated by the equation  $P_0 = e^{-m}$ , where  $m$  is  $L/n$ , the fold coverage." For instance, for a genome of 2.8 Mb,  $m=1$  when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point,  $P_0 = e^{-1} = 0.37$ . The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence  $L$  has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size  $L$ , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length,  $G$ , is determined by the equation  $G = Le^{-m}$ , and the average gap size,  $g$ , follows the equation,  $g = L/n$ . Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

**2. Random Library Construction**

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

*Staphylococcus aureus* DNA was prepared by phenol extraction. A mixture containing 600 ug DNA in 3.3 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 30% glycerol was sonicated for 1 min. at 0°C in a Branson Model 450 Sonicator at the lowest energy setting using a 3 mm probe. The sonicated DNA was ethanol precipitated and redissolved in 500 ul TE buffer.

To create blunt-ends, a 100 ul aliquot of the resuspended DNA was digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200 ul BAL31 buffer. The digested DNA was phenol-extracted, ethanol-precipitated, redissolved in 100 ul TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size was excised from the gel, and the LGT agarose was melted and the resulting solution was extracted with phenol to separate the agarose from the DNA. DNA was ethanol precipitated and redissolved in 20 ul of TE buffer for ligation to vector.

A two-step ligation procedure was used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50 ul) contained 2 ug of DNA fragments, 2 ug pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and was incubated at 14°C for 4 hr. The ligation mixture then was phenol extracted and ethanol precipitated, and the precipitated DNA was dissolved in 20 ul TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder were visualized by ethidium bromide-staining and UV illumination and identified by size as insert (i), vector (v), v+i, v+2i, v+3i, etc. The portion of the gel containing v+i DNA was excised and the v+i DNA was recovered and resuspended into 20 ul TE. The v+i DNA then was blunt-ended by T4 polymerase treatment for 5 min. at 37° C in a reaction mixture (50 ul) containing the v+i linears, 500 uM each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+i linears were dissolved in 20 ul TE. The final ligation to produce circles was carried out in a 50 ul reaction containing 5 ul of v+i linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture was stored at -20°C.

This two-stage procedure resulted in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E.coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies* 3 (1):5 (1990)) were used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells were plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating was carried out as follows. A 100 ul aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) was thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7 ul aliquot of 1.42 M beta-

mercaptoethanol was added to the aliquot of cells to a final concentration of 25 mM. Cells were incubated on ice for 10 min. A 1 µl aliquot of the final ligation was added to the cells and incubated on ice for 30 min. The cells were heat pulsed for 30 sec. at 42° C and placed back on ice for 2 min. The outgrowth period in liquid culture was eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture was plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl<sub>2</sub> (1 M), and 1 ml MgSO<sub>4</sub>/100 ml SOB agar. The 15 ml top layer was poured just prior to plating. Our titer was approximately 100 colonies/10 µl aliquot of transformation.

All colonies were picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products would be deleted from the library, resulting in a slight increase in gap number over that expected.

### 3. Random DNA Sequencing

High quality double stranded DNA plasmid templates were prepared using an alkaline lysis method developed in collaboration with 5Prime → 3Prime Inc. (Boulder, CO). Plasmid preparation was performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Average template concentration was determined by running 25% of the samples on an agarose gel. DNA concentrations were not adjusted.

Templates were also prepared from a *Staphylococcus aureus* lambda genomic library. An unamplified library was constructed in Lambda DASH II vector (Stratagene). *Staphylococcus aureus* DNA (> 100 kb) was partially digested in a reaction mixture (200 µl) containing 50 µg DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23 C. The digested DNA was phenol-extracted and centrifuges over a 10- 40% sucrose gradient. Fractions containing genomic DNA of 15-25 kb were recovered by precipitation. One µl of fragments was used with 1 µl of DASHII vector (Stratagene) in the recommended ligation reaction. One µl of the ligation mixture was used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract Phage were plated directly without amplification from the packaging mixture (after dilution with 500 µl of recommended SM buffer and chloroform treatment). Yield was about 2.5x10<sup>9</sup> pfu/µl.

An amplified library was prepared from the primary packaging mixture according to the manufacturer's protocol. The amplified library is stored frozen in 7% dimethylsulfoxide. The phage titer is approximately 1x10<sup>9</sup> pfu/ml.

Mini-liquid lysates (0.1µl) are prepared from randomly selected plaques and template is prepared by long range PCR. Samples are PCR amplified using modified T3 and T7 primers, and Elongase Supermix (LTI).

Sequencing reactions are carried out on plasmid templates using a combination of two workstations (BIOMEK 1000 and Hamilton Microlab 2200) and the Perkin-Elmer 9600 thermocycler with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers. Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. Modified T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are on a combination of AB 373 DNA Sequencers and ABI 377 DNA sequencers. All of the dye terminator sequencing reactions are analyzed using the 2X 9 hour module on the AB 377. Dye primer reactions are analyzed on a combination of ABI 373 and ABI 377 DNA sequencers. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

### 4. Protocol for Automated Cycle Sequencing

The sequencing was carried out using Hamilton Microstation 2200, Perkin Elmer 9600 thermocyclers, ABI 373 and ABI 377 Automated DNA Sequencers. The Hamilton combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates were combined in the wells of a 96-well thermocycling plate and transferred to the Perkin Elmer 9600 thermocycler. Thirty consecutive cycles of linear amplification (i.e., one primer synthesis) steps were performed including denaturation, annealing of primer and template, and extension; i.e., DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols were used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer was labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 or 377 DNA Sequencer for electrophoresis, detection, and base-

calling. ABI currently supplies premixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions were loaded per ABI 373 Sequencer each day and 96 samples can be loaded on an ABI 377 per day. Electrophoresis was run overnight (ABI 373) or for 2 1/2 hours (ABI 377) following the manufacturer's protocols. Following electrophoresis and fluorescence detection, the ABI 373 or ABI 377 performs automatic lane tracking and base-calling. The lane-tracking was confirmed visually. Each sequence electropherogram (or fluorescence lane trace) was inspected visually and assessed for quality. Trailing sequences of low quality were removed and the sequence itself was loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence was removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 or ABI 377 were around 400 bp and depend mostly on the quality of the template used for the sequencing reaction.

## INFORMATICS

### 1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi-user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

### 2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than  $10^4$  fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164: 765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain ranged of base pairs (definable for each clone based on the known clone size range for a given library).

### 3. Identifying Genes

The predicted coding regions of the *Staphylococcus aureus* genome were initially defined with the program zorf, which finds ORFs of a minimum length. The predicted coding region sequences were used in searches against a database of all *Staphylococcus aureus* nucleotide sequences from GenBank (release 92.0), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs of at least 80 amino acids that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases.

ORFs of at least 120 amino acids that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.

## ILLUSTRATIVE APPLICATIONS

### 1. Production of an Antibody to a *Staphylococcus aureus* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

### 2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., Meth. Enzymol. 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.* Basic Methods in Molecular Biology Elsevier, New York. Section 21-2 (1989).

### 3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, J. Clin. Endocrinol. Metab. 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: Handbook of Experimental Immunology, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980).

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, they are useful in various animal models of Staphylococcal disease known to those of skill in the art as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic reagent.

### 3. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Staphylococcus aureus* genome, such as those of Tables 1-3 and SEQ ID NOS:1-5,191 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approxi-



mately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

#### 4. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Staphylococcus aureus* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Staphylococcus aureus* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Staphylococcus aureus* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Staphylococcus aureus* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Staphylococcus aureus* DNA 3' primer, taking care to ensure that the *Staphylococcus aureus* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Staphylococcus aureus* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Staphylococcus aureus* DNA.

Alternatively and if antibody production is not possible, the *Staphylococcus aureus* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Staphylococcus aureus* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* Express™ Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
1	1	1419	757	[emb]M17301 SAND	S. aureus DNA for hld gene and for part of agr gene	100	563	663
1	2	3273	2452	[emb]K52343 SAAG	S. aureus agrA, agrB and hld genes	99	809	822
1	5	6418	5651	[dbj]D14711 STAM	Staphylococcus aureus HSP10 and HSP60 genes	98	223	766
5	1	807	439	[emb]X72700 SAPV	S. aureus genes for S and P components of Pencon-Valentine leucocidins	81	216	369
5	4	3031	3571	[emb]X72700 SAPV	S. aureus genes for S and P components of Pencon-Valentine leucocidins	95	424	1461
10	1	86	904	[gb]L25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (gria and grib) genes, complete cds	98	715	819
16	5	5302	6246	[gb]U35773	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	94	251	945
16	6	6249	7091	[gb]U35773	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	99	843	843
16	7	7084	7584	[gb]U35773	Staphylococcus aureus prolipoprotein diacylglyceryl transferase (lgt) gene, complete cds	99	342	501
20	1	995	549	[gb]U19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	443	447
20	2	1011	841	[gb]U19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	91	137	171
20	3	2010	1798	[gb]U19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	100	110	213
20	4	5300	3825	[gb]M76714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	948	1476
20	5	4788	4282	[gb]M76714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	309	507
26	1	2	145	[gb]U41072	Staphylococcus aureus isoleucyl-tRNA synthetase (iles) gene, partial cds	100	126	144
26	2	84	557	[gb]U41072	Staphylococcus aureus isoleucyl-tRNA synthetase (iles) gene, partial cds	99	430	474
26	3	763	3531	[emb]X74219 SAIL	S. aureus gene for isoleucyl-tRNA synthetase	99	2769	2769
29	3	1261	4392	[gb]U66665	Staphylococcus aureus DNA fragment with class II promoter activity	100	117	3132
31	14	14977	13463	[emb]X73889 SAP1	S. aureus genes p1 and p2	99	1351	1515
31	15	14241	13855	[emb]X73889 SAP1	S. aureus genes p1 and p2	98	258	387
36	17	14284	13112	[gb]M12715	S. aureus gdh gene encoding lipase (glycerol ester hydrolase)	100	372	1173
36	19	13634	15518	[gb]M12715	S. aureus gdh gene encoding lipase (glycerol ester hydrolase)	100	2085	2085

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
46	2	519	1727	[gb U73374]	Staphylococcus aureus type 8 capsulae genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1209	1209
46	3	1720	2395	[gb U73374]	Staphylococcus aureus type 8 capsulae genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	576	576
46	4	2259	3182	[gb U73374]	Staphylococcus aureus type 8 capsulae genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	924	924
46	5	3173	4498	[gb U73374]	Staphylococcus aureus type 8 capsulae genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1283	1326
46	6	4536	5720	[gb U73374]	Staphylococcus aureus type 8 capsulae genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1185	1185
46	7	6455	6120	[gb U73374]	Staphylococcus aureus type 8 capsulae genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	278	336
48	1	2	955	[gb L25693]	Staphylococcus aureus recA gene, complete cds	99	954	954
50	3	4465	2924	[emb X85029 SAAH]	S. aureus AhpC gene	100	88	1542
50	4	4108	3315	[emb X85029 SAAH]	S. aureus AhpC gene	98	540	594
54	3	5074	3392	[emb X62992 SAFN]	S. aureus fnbB gene for fibronectin binding protein B	100	1668	1683
54	4	4865	4122	[emb X62992 SAFN]	S. aureus fnbB gene for fibronectin binding protein B	99	720	744
54	5	5056	4562	[emb X62992 SAFN]	S. aureus fnbB gene for fibronectin binding protein B	100	463	495
54	6	11386	8300	[gb J04151]	S. aureus fibronectin-binding protein (fnbA) mRNA, complete cds	100	3087	3087
56	3	1743	2819	[emb X87104 SADN]	S. aureus mdr, pbp4 and teqD genes (SG311-55 isolate)	89	68	1077
58	4	2858	3280	[emb X91786 SAPB]	S. aureus abcA, pbp4, and tagD genes	99	423	423
58	5	6005	4701	[emb X91786 SAPB]	S. aureus abcA, pbp4, and tagD genes	99	1305	1305
58	6	5677	5378	[gb U29478]	Staphylococcus aureus ABC transporter-like protein AbcA (abcA) gene, partial cds	100	300	300
58	7	5086	6840	[emb X91786 SAPB]	S. aureus abcA, pbp4, and tagD genes	99	1755	1755
72	1	888	445	[gb U21854]	S. aureus agr gene encoding an accessory gene regulator protein, complete cds	100	444	444
72	2	2457	1453	[emb X52543 SAAG]	S. aureus agrA, agrB and hld genes	99	673	1005

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	HSP nt length	ORF nt length
82	1	357	3937	emb X64172 SARP	S.aureus rplL, orf202, rpoB1r1 and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF203, DNA-directed RNA polymerase beta & beta' chains	99	2396	3561
82	2	4027	7677	emb X89233 SARP	S.aureus DNA for rpoC gene	99	3171	3651
82	3	7745	8068	gb U20869	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	320	324
82	4	8103	8579	gb U20869	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	477	477
82	5	8618	8821	gb U20869	Staphylococcus aureus ribosomal protein S12 (rpsL) gene, complete cds, ribosomal protein S7 (rpsG) and ORF 1 genes, partial cds	100	154	204
84	1	18	191	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	166	174
84	2	189	893	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	94	705	705
84	3	887	1660	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	774	774
84	4	1584	3503	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	98	1920	1920
84	5	3394	4521	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1128	1128
84	6	4519	5643	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	97	1125	1125
96	2	1245	3896	emb 218052 SACP	S.aureus gene for clumping factor	83	660	2652
97	2	625	882	gb U41072	Staphylococcus aureus isoleucyl-tRNA synthetase (ileS) gene, partial cds	97	68	258
111	1	3	452	gb U41499	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	450	450
111	2	526	1041	gb U41499	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	516	516
117	2	1278	1958	gb M83994	Staphylococcus aureus prolipoprotein signal peptidase (lep) gene, complete cds	100	61	681

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
118	4	3787	4254	[dbj U0690 STAN	Staphylococcus aureus genes for ORF37: HSP20; HSP70; HSP40: ORF35, complete cds	99	467	468
130	4	2597	3640	[emb X13290 SATW	Staphylococcus aureus multi-resistance plasmid pSK1 DNA containing tetracycline Tn4003	78	956	1044
130	5	3813	4265	[emb Z16422 SADI	S. aureus dfrB gene for dihydrofolate reductase	98	416	453
130	6	4309	5172	[emb Z16422 SADI	S. aureus dfrB gene for dihydrofolate reductase	98	607	864
136	4	5296	6207	[emb K71437 SAGY	S. aureus genes gyrB, gyrA and recF (partial)	97	838	912
136	5	11680	8987	[dbj U0489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	2694	2694
136	6	12886	10940	[dbj U0489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	1947	1947
136	7	12592	11765	[gb S77055	recF cluster: dnaA-replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 3573 nt]	99	822	828
143	3	4171	2867	[gb U36379	Staphylococcus aureus 5-adenosylmethionine synthetase gene, complete cds	99	1305	1305
143	4	3100	4281	[gb L42943	Staphylococcus aureus (clone KIN50) phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	1170	1182
143	5	4254	4718	[gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxykinase (pckA) gene, complete cds	100	449	465
143	9	6977	7261	[gb U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (ene), and o-succinylbenzoic acid synthetase (henc) genes, complete cds	100	75	285
143	10	9464	8361	[gb U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (ene), and o-succinylbenzoic acid synthetase (henc) genes, complete cds	100	1104	1104
143	11	11232	9768	[gb U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (ene), and o-succinylbenzoic acid synthetase (henc) genes, complete cds	100	1485	1485
143	12	10739	10320	[gb U51132	Staphylococcus aureus o-succinylbenzoic acid CoA ligase (ene), and o-succinylbenzoic acid synthetase (henc) genes, complete cds	100	332	420
152	5	2454	3437	[emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	305	984
152	6	3513	4820	[emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	98	1308	1308
152	7	4818	6230	[emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	1413	1413
153	1	387	1526	[gb S77055	recF cluster: dnaA-replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 3573 nt]	99	1140	1140
153	2	1877	2152	[gb S77055	recF cluster: dnaA-replisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 3573 nt]	100	276	276

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
153	3	2143	2289	[gb S77055]	lecF cluster: dihaereplisome assembly protein...gyrB-DNA gyrase beta subunit [Staphylococcus aureus, Y8886, Genomic, 5 genes, 3573 nt]	99	113	147
154	10	10792	9314	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	91	154	1479
154	11	9935	9615	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	99	229	321
154	12	9943	10167	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	94	123	225
154	13	10089	11501	[gb U06451]	[Staphylococcus aureus proline permease homolog (putP) gene, complete cds]	99	1326	1413
159	2	2195	1212	[dbj D28879]	[Staphylococcus aureus gene for penicillin-binding protein 1, complete cds]	100	71	984
161	3	2596	2270	[gb M83994]	[Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds]	92	203	327
162	1	1406	705	[gb U21221]	[Staphylococcus aureus hyaluronate lyase (hysA) gene, complete cds]	100	702	702
163	4	1263	1772	[gb U19770]	[Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds]	96	127	510
164	7	4774	9117	[dbj D86727]	[Staphylococcus aureus DNA for DNA polymerase III, complete cds]	99	3470	4344
168	7	7448	6447	[gb U21636]	[Staphylococcus aureus cnp-binding-factor 1 (cbf1) and ORF X genes, complete cds]	100	1002	1002
168	8	9538	7961	[gb U21636]	[Staphylococcus aureus cnp-binding-factor 1 (cbf1) and ORF X genes, complete cds]	99	1158	1578
173	6	9240	7801	[gb J03479]	[S.aureus enzyme II-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	1440	1440
173	7	11252	9522	[gb J03479]	[S.aureus enzyme II-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	99	1731	1731
173	8	8285	6704	[gb J03479]	[S.aureus enzyme II-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	420	420
173	9	10168	9839	[gb J03479]	[S.aureus enzyme II-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds]	100	330	330
173	10	111815	10829	[emb X14827]	[SALSA] [Staphylococcus aureus lacC and lacD genes]	100	987	987
173	11	112721	11774	[emb X14827]	[SALSA] [Staphylococcus aureus lacC and lacD genes]	100	948	948
173	12	112838	12305	[gb M64724]	[S.aureus tagatase 6-phosphate isomerase gene, complete cds]	100	534	534
173	13	112243	12773	[gb M32103]	[Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds]	100	471	471
173	14	114633	13866	[gb M32103]	[Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds]	100	768	768

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
176	1	2	655	[gb U52961]	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	115	654
178	2	2201	1482	[gb U52961]	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	720	720
178	3	2361	1909	[gb U52961]	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	453	453
178	4	1551	1853	[gb U52961]	Staphylococcus aureus holin-like protein LrgA (lrgA) and LrgB (lrgB) genes, complete cds	100	303	303
178	5	3541	2777	[gb L42945]	Staphylococcus aureus lytS and lytR genes, complete cds	99	765	765
178	6	3294	3025	[gb L42945]	Staphylococcus aureus lytS and lytR genes, complete cds	99	270	270
181	1	1114	590	[gb H6177]	S. aureus sigma factor (plac) gene, complete cds	99	499	525
182	1	3	341	[emb X61307 SASP]	Staphylococcus aureus spa gene for protein A	98	277	339
182	2	690	2312	[gb J01786]	S. aureus spa gene coding for protein A, complete cds	97	1332	1623
182	3	5861	4251	[emb X61307 SASP]	Staphylococcus aureus spa gene for protein A	99	119	1611
185	1	3	824	[gb U31979]	Staphylococcus aureus chorismate synthase (arcC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	90	132	822
191	3	841	2760	[emb X17679 SACO]	Staphylococcus aureus coa gene for coagulase	99	1920	1920
191	4	2967	3143	[emb X16457 SAST]	Staphylococcus aureus gene for staphylocoagulase	99	177	177
191	5	5768	4566	[emb X16457 SAST]	Staphylococcus aureus gene for staphylocoagulase	99	250	1203
196	1	1741	872	[gb L36472]	Staphylococcus aureus lysyl-tRNA synthetase gene, complete cds, transfer RNA (tRNA) genes, 5S ribosomal RNA (5S rRNA) gene, 16S ribosomal RNA (16S rRNA) gene, 23S ribosomal RNA (23S rRNA) gene	99	870	870
198	3	1688	2011	[emb X91205 SAPT]	S. aureus ptaiH and ptaiI genes	99	324	324
198	4	2005	2310	[emb X91205 SAPT]	S. aureus ptaiH and ptaiI genes	97	304	306
202	1	163	1305	[emb X97985 SA12]	S. aureus orf1, 2, 3 & 4	99	1143	1143
202	2	1303	2175	[emb X71889 SAP1]	S. aureus genes P1 and P2	94	464	873
210	1	3114	1558	[dbj D17366 STAA]	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1552	1557
210	2	2939	2232	[gb L41499]	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	99	684	708
214	11	7439	7770	[dbj D86240 D862]	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	96	157	342

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
216	3	398	1318	emb X72700 SAPV	S.aureus genes for S and F components of Pantone-Valentine leucocidins	88	265	921
219	2	1810	1073	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	100	60	738
219	3	2979	2035	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	99	945	945
219	4	4359	3196	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	99	1164	1164
219	5	7044	1166	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	98	1869	1869
219	6	6557	5883	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	99	675	675
219	7	6801	6314	dbj D30690 STAN	Staphylococcus aureus genes for ORF37; HSP70; HSP40; ORF35, complete cds	98	468	468
221	8	10816	10014	gb U19298	Staphylococcus aureus phosphatidylinositol-specific phospholipase C (pic) gene, complete cds	91	67	783
223	1	2855	1506	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	102	1350
234	1	2	1357	emb X97985 SA12	S.aureus orf1.2.3 & 4	100	176	1356
234	2	1694	2485	emb X97985 SA12	S.aureus orf1.2.3 & 4	100	792	792
234	3	2648	3148	emb X97985 SA12	S.aureus orf1.2.3 & 4	99	501	501
234	4	3120	4604	emb X97985 SA12	S.aureus orf1.2.3 & 4	99	1305	1485
236	6	3826	5322	gb U48826	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	96	648	1497
248	1	2	403	emb X62288 SAP2	S.aureus DNA for penicillin-binding protein 2	100	103	402
248	2	388	852	gb U35426	Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	465	465
253	2	1519	1093	gb U46541	Staphylococcus aureus sara gene, complete cds	96	447	447
254	2	150	1835	gb U57060	Staphylococcus aureus scdA gene, complete cds	94	142	1686
254	3	1973	2728	gb U57060	Staphylococcus aureus scdA gene, complete cds	99	756	756
260	1	2	1900	gb U90653	Staphylococcus aureus glycerol ester hydrolase (lip) gene, complete cds	99	1213	1899
265	1	1	942	dbj D21131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	941	942

TABLE 1



S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent Ident	ORF nt length
265	2	688	476	dbj D21131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	99	213
265	3	2418	1765	dbj D21131 STAS	Staphylococcus aureus gene for a participant in homogeneous expression of high-level methicillin resistance, complete cds	98	654
266	1	2	1018	dbj D14711 STAH	Staphylococcus aureus MSP10 and MSP60 genes	98	743
282	1	1	525	gb F72488	hemA-porphobilinogen synthase [Staphylococcus aureus, SA1959, Genomic, 1087 nt]	100	110
282	2	516	1502	gb S72488	hemB-porphobilinogen synthase [Staphylococcus aureus, SA1959, Genomic, 1087 nt]	100	952
284	1	3	170	gb H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	84
284	2	282	1034	gb H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	712
284	3	1028	2026	gb H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	979
284	4	1990	2202	gb H63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	98	187
289	3	1536	1991	gb H32470	S. aureus SauIAI-restriction-enzyme and SauIAI-modification-enzyme genes, complete cds	99	338
303	1	2	868	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (higa, hlgB, hlgC) genes, complete cds	99	867
303	2	1409	2383	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (higa, hlgB, hlgC) genes, complete cds	100	975
303	3	2367	3161	gb L01055	Staphylococcus aureus gamma-hemolysin components A, B and C (higa, hlgB, hlgC) genes, complete cds	99	793
305	1	2707	1355	dbj D17366 STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	1343
311	1	2628	1335	gb L42945	Staphylococcus aureus lytS and lytR genes, complete cds	98	1314
312	6	7019	7870	gb L14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	74	351
323	1	1998	1003	gb U31175	Staphylococcus aureus D-specific D-2-hydroxyacid dehydrogenase (ddh) gene, complete cds	98	996
326	1	1	237	emb Y00356 SASP	Staphylococcus aureus V8 serine protease gene	100	108
338	1	687	188	emb X64189 SALE	S. aureus leuP-P83 gene for F component of leucocidin R	98	259
338	2	1828	1088	emb X64189 SALE	S. aureus leuP-P83 gene for F component of leucocidin R	97	137

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
342	2	579	1754	gb U06462	Staphylococcus aureus S44 FtaZ (ftsZ) gene, complete cds	100	1176	1176
344	2	517	1248	emb V01281 SARU	S. aureus mRNA for nuclease	98	732	732
349	1	457	230	gb M20393	S. aureus bacteriophage phi-11 attachment site (attB)	95	172	228
353	1	1016	516	gb M3994	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	100	187	501
353	2	1582	1046	gb M83994	Staphylococcus aureus prolipoprotein signal peptidase (lsp) gene, complete cds	99	537	537
356	1	3	674	gb U20503	Staphylococcus aureus MHC class II analog gene, complete cds	75	671	672
361	1	1	903	gb L19298	Staphylococcus aureus phosphatidylinositol-specific phospholipase C (pic) gene, complete cds	98	747	903
361	2	1103	1507	gb L19298	Staphylococcus aureus phosphatidylinositol-specific phospholipase C (pic) gene, complete cds	97	68	405
373	1	3	1148	emb M62288 SAFE	S. aureus DNA for penicillin-binding protein 2	99	1146	1146
389	3	1904	1248	emb M62282 SATS	S. aureus target site DNA for IS431 insertion	97	349	657
400	1	1	540	emb M61716 SAHL	S. aureus hlb gene encoding sphingomyelinase	99	389	540
400	2	1693	1187	emb M13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	99	178	507
408	1	1810	1049	gb S76213	asp23-alkaline shock protein 23 (methicillin resistant) (Staphylococcus aureus, 912, Genomic, 1360 nt)	99	163	762
418	1	2	217	gb L41499	Staphylococcus aureus ORF1, partial cds, ORF2, ORF3, autolysin (atl) genes, complete cds	100	216	216
418	2	854	639	dbj D17366 STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	100	188	216
421	2	1262	2509	gb L43098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	1248	1248
422	1	2	325	gb K02985	S. aureus (strain RN450) transposon Tn554 insertion site	96	200	374
427	1	865	434	dbj D28879 STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	432	432
427	2	1829	1122	dbj D28879 STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	100	151	708
435	1	2	808	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	100	556	807
435	2	832	999	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	100	134	168
436	1	1341	685	emb M17688 SAFE	S. aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end	97	657	657

TABLE 1

## TABLE I

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
436	2	2403	1657	[emb N17668 SAFE	S.aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' and	100	294	747
442	1	1347	1300	[emb X72700 SAVP	S.aureus genes for S and F components of Panton-Valentine leucocidins	84	204	954
445	2	1906	2178	[gb U01055	Staphylococcus aureus gamma-hemolysin components A, B and C (higA, hlgB, hlgC) genes, complete cds	98	187	273
447	1	167	1078	[gb U13770	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	100	51'	912
447	2	1176	1784	[gb U13770	Staphylococcus aureus pyrrolidone carboxyl peptidase (pcp) gene, complete cds	96	597	609
454	3	7309	4319	[emb 218852 SACF	S.aureus gene for clumping factor	75	653	2991
472	4	7896	5479	[gb U25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grIA and grIB) genes, complete cds	99	2418	2418
472	5	8120	6792	[gb U25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grIA and grIB) genes, complete cds	99	1328	1329
475	2	566	889	[emb X52543 SAAQ	S.aureus agrA, agrB and hld genes	100	76	324
481	4	1922	1560	[emb X64172 SARP	S.aureus rplL, rpf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	250	363
481	5	1244	1534	[emb X64172 SARP	S.aureus rplL, rpf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	224	291
487	2	1388	1188	[gb U83994	Staphylococcus aureus prolipoprotein signal peptidase (lpp) gene, complete cds	98	72	201
489	1	2737	1370	[gb U21221	Staphylococcus aureus hyaluronate lyase (hyaA) gene, complete cds	99	1368	1368
503	2	1135	653	[gb U83994	Staphylococcus aureus prolipoprotein signal peptidase (lpp) gene, complete cds	100	108	483
511	3	1613	2242	[gu L14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	323	630
511	4	3122	2700	[gu S76213	asp23-saikaline shock protein 23 (methicillin resistant) [Staphylococcus aureus, 912, Ganomic, 1360 nt]	96	423	423
520	2	758	1297	[emb X72014 SAFI	S.aureus fib gene for fibrinogen-binding protein	99	540	540
520	3	1436	1801	[emb X72013 SAFI	S.aureus fib gene for fibrinogen-binding protein	99	221	366
526	1	2150	1092	[dbj U17366 STAA	Staphylococcus aureus atl gene for autolysin, complete cds and other ORFs	99	641	1059

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	HSP nt length	ORF nt length
528	2	58	963	gb U19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homolog, 5' flank	99	260	906
528	3	1098	2870	gb U19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homolog, 5' flank	99	866	1773
530	1	3	434	gb U31979	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds; dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	99	432	432
530	2	1211	2395	gb U31979	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds; dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	91	1185	1185
530	3	2409	2801	gb U31979	Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds; dehydroquinase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	88	181	393
530	4	2690	3484	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	75	795
530	5	3482	4792	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	99	905	1311
530	6	4790	5380	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	196	591
539	1	3	338	emb X76490 SAGL	S. aureus (bb270) glxA and glxR genes	99	336	336
539	2	336	527	emb X76490 SAGL	S. aureus (bb270) glxA and glxR genes	100	189	192
534	1	727	365	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	100	54	363
534	2	2175	1252	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	918	924
534	3	1574	1374	gb U73374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	96	122	201
584	2	1019	705	gb U21221	Staphylococcus aureus hyaluronate lyase (hyla) gene, complete cds	99	306	315
587	3	1475	4288	emb 218852 SACF	S. aureus gene for clumping factor	98	2588	2814
598	1	3881	1953	dbj D28879 STAP	Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	99	1873	1929

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
605	1	2	745	U0106240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	98	338	744
609	1	1628	816	emb X176490 SAGL	S. aureus (bb270) glpA and glpR genes	100	495	813
614	1	1280	642	gb U12103	Staphylococcus aureus lac repressor (lacR) gene, complete cds and lacA repressor (lacA), partial cds	99	639	639
626	1	2508	1255	gb U63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	100	225	1254
626	2	3315	2284	gb U63176	Staphylococcus aureus helicase required for T181 replication (pcrA) gene, complete cds	99	838	1032
629	1	1999	1001	emb X17688 SAFE	S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	99	990	999
629	2	1407	1195	emb X17688 SAFE	S. aureus factor essential for expression of methicillin resistance (femA) gene, complete cds, and trpA gene, 3' end	98	194	213
631	2	3126	3228	emb Z18852 SACP	S. aureus gene for clumping factor	82	489	1899
632	1	3	551	emb Z10588 SAST	S. aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	549	549
632	2	529	1323	emb Z10588 SAST	S. aureus (RN4220) genes for potential ABC transporter and potential membrane spanning protein	99	795	795
651	1	1909	1070	gb U19300	Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank	99	478	840
657	2	1800	1105	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	84	456	696
662	1	908	456	emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	359	453
662	2	230	475	emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	100	246	246
662	3	746	1399	emb X13404 SAHL	Staphylococcus aureus hlb gene for beta-hemolysin	99	653	654
682	1	956	480	gb U63177	S. aureus sigma factor (placI) gene, complete cds	100	136	477
685	1	1182	592	gb U65000	Staphylococcus aureus type-I signal peptidase SpSA (spSA) gene, and type-I signal peptidase SpSB (spSB) gene, complete cds	98	514	591
685	2	1716	1153	gb U65000	Staphylococcus aureus type-I signal peptidase SpSA (spSA) gene, and type-I signal peptidase SpSB (spSB) gene, complete cds	96	564	564
697	1	3	527	gb U63177	S. aureus sigma factor (placI) gene, complete cds	100	195	525
697	2	485	784	gb U63177	S. aureus sigma factor (placI) gene, complete cds	97	280	300

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent Ident	HSP nt length	ORF nt length
710	1	15	503	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	217	489
733	1	26	205	gb M80252	Staphylococcus aureus norA1199 gene (which mediates active efflux of fluoroquinolones), complete cds	97	160	180
741	1	1736	1197	dbj U83951 STAL	Staphylococcus aureus DNA for LukM component, LukP-PV like component, complete cds	81	522	540
752	1	1	636	emb Y00036 SASP	Staphylococcus aureus V8 serine protease gene	99	618	636
752	2	588	936	emb Y00036 SASP	Staphylococcus aureus V8 serine protease gene	99	340	369
756	1	1308	709	emb X01645 SATO	Staphylococcus aureus (Wood 46) gene for alpha-toxin	98	567	600
777	1	1582	950	emb Z49245 SA42	S. aureus partial sod gene for superoxide dismutase	99	429	633
780	1	1113	537	gb U20503	Staphylococcus aureus NHC class-II analog gene, complete cds	86	550	555
784	1	73	687	gb U63329	Staphylococcus aureus novel antigen gene, complete cds	99	568	615
797	1	182	544	dbj D14711 STAN	Staphylococcus aureus HSP10 and HSP60 genes	98	363	363
798	1	532	302	emb X58434 SNPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	95	196	231
823	1	3	467	gb S77055	ref cluster: dnaArepilosome assembly protein...gyrB-dna gyrase beta subunit [Staphylococcus aureus, Y886, Genomic, 5 genes, 3573 nt]	99	156	465
848	1	348	175	gb L25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	99	174	174
848	2	476	318	gb L25288	Staphylococcus aureus gyrase-like protein alpha and beta subunit (grlA and grlB) genes, complete cds	100	131	159
866	1	792	397	emb X64172 SARP	S. aureus rplL, orf202, rpoB(r1) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	395	396
883	1	1	285	dbj D90119 STAM	S. aureus norA gene	99	131	285
884	1	606	334	emb X52543 SAG	S. aureus agrA, agrB and hid genes	98	265	273
884	2	716	522	emb X52543 SAG	S. aureus agrA, agrB and hid genes	100	195	195
912	2	517	681	emb Z30588 SAST	S. aureus (RM4220) genes for potential ABC transporter and potential membrane spanning protein	99	163	165
917	1	2	265	gb M64724	S. aureus tagatase 6-phosphate isomerase gene, complete cds	99	247	264
917	2	238	396	gb M64724	S. aureus tagatase 6-phosphate isomerase gene, complete cds	95	147	159
918	1	2426	1215	emb X93205 SAPT	S. aureus ptmH and ptmI genes	99	1212	1212

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	percent ident	ORF nt length	ORF nt length
967	1	1	411	dbj U90119 STAM	S. aureus norA gene	97	395	411
991	1	672	337	emb X52543 SMAG	S. aureus agrA, agrB and hid genes	99	336	336
1000	1	1117	845	gb L14017	Staphylococcus aureus methicillin-resistance protein (mecA) gene and unknown ORF, complete cds	78	190	273
1001	1	498	265	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	234	234
1010	1	1	285	gb U21221	Staphylococcus aureus hyaluronate lyase (hlyA) gene, complete cds	99	224	285
1046	1	656	330	emb X72700 SAPV	S. aureus genes for S and F components of Panton-Valentine leucocidins	85	205	327
1060	1	480	286	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliponamide acetyltransferase and dihydroliponamide dehydrogenase	99	180	195
1073	1	1176	589	gb K02985	S. aureus (strain RN450) transposon Tn554 insertion site	100	131	588
1079	1	3	230	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	228	228
1079	2	218	484	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	267	267
1079	3	460	645	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	100	186	186
1092	1	289	146	emb X58434 SAPD	S. aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliponamide acetyltransferase and dihydroliponamide dehydrogenase	98	124	144
1143	1	1	243	gb H03177	S. aureus sigma factor (nlaC) gene, complete cds	99	243	243
1157	1	2	136	emb Z48003 SNON	S. aureus gene for DNA polymerase III	97	127	135
1189	1	720	361	gb S74031	norA-NorA (ISP794) [Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt]	99	360	360
1190	1	2	283	gb H21854	S. aureus agr gene encoding an accessory gene regulator protein, complete cds	100	282	282
1190	2	1127	888	emb X52543 SMAG	S. aureus agrA, agrB and hid genes	100	240	240
1225	1	2	163	emb X17679 SACD	Staphylococcus aureus coa gene for coagulase	97	124	162
1243	1	2	529	dbj D86240 D862	Staphylococcus aureus gene for unknown function and dlt operon dltA, dltB, dltC and dltD genes, complete cds	99	495	528
1244	1	1	210	gb S74031	norA-NorA (ISP794) [Staphylococcus aureus, NCTC 8325, Insertion, 1820 nt]	100	210	210
1301	1	41	472	emb X76490 SACB	S. aureus (bs270) glnA and glnB genes	99	299	432

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	HSP nt length	ORF nt length
1315	1	18	326	[emb]X64172 SARP	[S.aureus rplL, orf202, rpoB] and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta 4 beta' chains	98	277	309
1519	1	2	175	[dhj]D28879 STAP	[Staphylococcus aureus gene for penicillin-binding protein 1, complete cds	98	139	174
1663	1	1346	675	[dhj]D86240 D862	[Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	98	672	672
1797	1	644	324	[gb]U73374	[Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	99	331	321
1857	1	1	192	[gb]M90536	[Staphylococcus aureus alpha-hemolysin gene, 3' end	98	192	192
1921	1	2	181	[emb]X17688 SAFE	[S.aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end	100	180	180
1957	1	2	346	[gb]U60589	[Staphylococcus aureus novel antigen gene, complete cds	99	345	345
1988	1	1	402	[dhj]D86240 D862	[Staphylococcus aureus gene for unknown function and dit operon dltA, dltB, dltC and dltD genes, complete cds	100	402	402
2100	1	414	408	[gb]M63177	[S.aureus sigma factor (plac) gene, complete cds	99	207	207
2199	1	1	402	[gb]U66664	[Staphylococcus aureus DNA fragment with class II promoter activity	99	131	402
2537	1	308	156	[emb]X17688 SAFE	[S.aureus factor essential for expression of methicillin resistance (fema) gene, complete cds, and trpA gene, 3' end	99	153	153
2891	1	2	400	[gb]U25426	[Staphylococcus aureus penicillin-binding protein 2 (pbp2) gene, complete cds	99	399	399
2950	1	778	398	[dhj]D30690 STAN	[Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	100	356	381
2971	1	3	398	[gb]U51132	[Staphylococcus aureus o-succinylbenzoic acid CoA ligase (fema), and o-succinylbenzoic acid synthetase (femC) genes, complete cds	97	272	396
2978	1	618	328	[gb]U31979	[Staphylococcus aureus chorismate synthase (aroC) and nucleoside diphosphate kinase (ndk) genes, complete cds, dehydrogenase synthase (aroB) and geranylgeranyl pyrophosphate synthetase homolog (gerCC) genes, partial cds	98	250	291
2985	1	832	464	[emb]X17679 SACO	[Staphylococcus aureus coa gene for coagulase	98	367	369
3006	1	2170	1784	[gb]U11779	[Staphylococcus aureus methicillin-resistant ATCC 33932 clone RNW30 16S-23S rRNA spacer region	87	82	387
3008	1	474	238	[dhj]D30690 STAN	[Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	88	178	237
3008	2	451	281	[dhj]U30690 STAN	[Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	97	120	171

TABLE 1



S aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP length	ORF nt length
3011	1	793	398	emb X62992 SAFN	S aureus fnbB gene for fibronectin binding protein B	93	72	396
3019	1	2	235	gb J03479	S aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	234	234
3023	1	81	233	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	87	100	153
3029	1	90	287	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	100	135	198
3039	1	18	164	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	97	135	167
3039	2	70	327	gb U51133	Staphylococcus aureus phosphoenolpyruvate carboxylase (pckA) gene, complete cds	77	183	258
3056	1	3	215	emb X64172 SARF	S aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	213	213
3059	1	1	261	dbj U0690 STAN	Staphylococcus aureus genes for ORF37; HSP20; HSP70; HSP40; ORF35, complete cds	98	234	261
3073	1	27	284	gb U06451	Staphylococcus aureus proline permease homolog (putP) gene, complete cds	99	229	258
3074	1	2	397	emb X64172 SARF	S aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	96	250	396
3088	1	3	239	dbj D8672 D867	Staphylococcus aureus DNA for DNA polymerase III, complete cds	95	215	237
3097	1	444	244	emb Z18003 SAHJ	S aureus gene for DNA polymerase III	97	160	201
3102	1	307	355	gb J03479	S aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	97	142	153
3121	1	568	398	emb X58434 SAPD	S aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	100	88	171
3125	1	463	233	emb X89233 SARF	S aureus DNA for rpoC gene	98	192	231
3133	1	2	375	emb Z18852 SACF	S aureus gene for clumping factor	96	154	174
3160	1	420	211	dbj U0489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	89	197	210
3176	1	1	378	emb X58434 SAPD	S aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	96	91	378
3192	1	420	211	gb J03479	S aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	72	210
3210	1	3	143	gb U76714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	96	141	141

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	Percent ident	ORF nt length
3332	3	2106	1282	[gb U14017]	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	71	257
3338	1	2	374	[emb X89331 SARP]	S. aureus DNA for rpoC gene	99	356
3443	1	392	634	[gb U11510]	Staphylococcus aureus transfer RNA sequence with two rRNAs	99	102
3555	1	637	320	[emb Z18852 SACF]	S. aureus gene for clumping factor	99	307
3559	1	3	182	[emb X17679 SACO]	Staphylococcus aureus coa gene for coagulase	100	143
3559	2	95	313	[emb X17679 SACO]	Staphylococcus aureus coa gene for coagulase	98	174
3563	1	278	141	[gb U35773]	Staphylococcus aureus prol::protein diacylglycerol transferase (lgt) gene, complete cds	100	79
3563	2	527	363	[gb U35773]	Staphylococcus aureus prolipoprotein diacylglycerol transferase (lgt) gene, complete cds	98	162
3566	1	3	422	[emb X16457 SAST]	Staphylococcus aureus gene for staphylocoagulase	98	175
3588	1	2	262	[gb U43098]	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	253
3593	1	1	350	[gb U03479]	S. aureus enzyme ltr-lac (lacF), enzyme ltr-lac (lacF), and phospho-beta-galactosidase (lacG) genes, complete cds	99	345
3600	1	758	381	[emb Z18852 SACF]	S. aureus gene for clumping factor	72	346
3602	1	788	396	[emb Z18852 SACF]	S. aureus gene for clumping factor	98	319
3656	1	1013	528	[emb Z18852 SACF]	S. aureus gene for clumping factor	84	403
3682	1	3	236	[emb X64172 SARP]	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	231
3682	2	224	415	[emb X64172 SARP]	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	112
3693	1	758	423	[emb X62992 SAFN]	S. aureus fnbB gene for fibronectin binding protein B	100	229
3702	1	593	354	[gb U11510]	Staphylococcus aureus transfer RNA sequence with two rRNAs	54	81
3725	1	924	463	[emb Z18852 SACF]	S. aureus gene for clumping factor	71	367
3761	1	809	450	[gb U14017]	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	85	333
3767	1	1	402	[emb X64172 SARP]	S. aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	387

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	ORF nt length
3775	1	2	266	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	227
3786	1	456	229	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	204
3786	2	542	366	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	95	123
3798	1	3	251	emb X17679 SACO	Staphylococcus aureus coa gene for coagulase	99	249
3813	1	793	398	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	396
3819	1	184	402	emb X68425 SA23	S.aureus gene for 23S rRNA	99	161
3844	1	932	468	gb U38826	Staphylococcus aureus elastin binding protein (ebpS) gene, complete cds	87	204
3845	1	1	381	emb X58434 SAPD	S.aureus pddB, pddC and pddD genes for pyruvate decarboxylase, dihydrolysoamide acetyltransferase and dihydrolysoamide dehydrogenase	94	356
3856	1	798	400	gb U44017	Staphylococcus aureus methicillin-resistance protein (mecA) gene and unknown ORF, complete cds	76	192
3859	1	1049	573	emb Z18852 SACF	S.aureus gene for clumping factor	85	347
3871	1	650	327	gb H76714	Staphylococcus aureus peptidoglycan hydrolase gene, complete cds	100	299
3876	1	2	253	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	100	217
3877	1	572	288	gb J03479	S.aureus enzyme III-lac (lacP), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacD) genes, complete cds	97	209
3878	1	1	237	emb X58434 SAPD	S.aureus pddB, pddC and pddD genes for pyruvate decarboxylase, dihydrolysoamide acetyltransferase and dihydrolysoamide dehydrogenase	96	155
3888	1	3	173	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	98	171
3893	1	1	183	emb X89233 SARP	S.aureus DNA for rpoC gene	100	170
3893	2	181	357	emb X89233 SARP	S.aureus DNA for rpoC gene	98	79
3894	1	3	485	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	450
3895	1	836	420	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	413
3905	1	48	239	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	100	159
3905	2	188	400	gb U05004	Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	97	88

TABLE I

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
3910	1	3	359	emb X58434 SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	278	357
3915	1	1	330	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	175	330
3964	1	691	347	emb Z48003 SADN	S.aureus gene for DNA polymerase III	100	295	345
4007	1	199	390	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	98	163	192
4036	1	3	371	dbj D10489 STAG	Staphylococcus aureus genes for DNA gyrase A and B, complete cds	99	339	369
4046	1	692	348	emb Z18852 SACF	S.aureus gene for clumping factor	87	221	345
4060	1	1	375	emb Z18852 SACF	S.aureus gene for clumping factor	96	271	375
4061	1	860	432	emb Z48003 SADN	S.aureus gene for DNA polymerase III	99	429	429
4062	1	606	304	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	198	303
4085	1	58	402	gb U11786	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RUW42 165-238 rRNA spacer region	98	127	345
4088	1	2	301	gb U143098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	99	227	300
4093	1	2	277	emb X58434 SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	99	276	276
4097	1	1	402	emb Z18852 SACF	S.aureus gene for clumping factor	74	307	402
4116	1	22	402	gb U05004	Staphylococcus aureus dehydroquininate synthase (aroB) gene, 3' end cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene, complete cds; ORF3, complete cds	98	157	381
4125	1	240	401	gb U073374	Staphylococcus aureus type 8 capsule genes, cap8A, cap8B, cap8C, cap8D, cap8E, cap8F, cap8G, cap8H, cap8I, cap8J, cap8K, cap8L, cap8M, cap8N, cap8O, cap8P, complete cds	100	86	162
4149	1	35	247	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	200	213
4151	1	629	366	gb U14017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	87	150	264
4154	1	754	398	emb X64172 SARP	S.aureus rplL, orf202, rpoB1r1f and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	297	357
4179	1	1	294	emb X64172 SARP	S.aureus rplL, orf202, rpoB1r1f and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	240	294

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
4203	1	1	255	emb X89233 SARP	S.aureus DNA for rpoC gene	99	239	255
4206	1	1	303	emb Z18052 SACF	S.aureus gene for clumping factor	100	236	303
4206	2	195	344	emb Z18052 SACF	S.aureus gene for clumping factor	95	65	150
4208	1	108	314	emb X58434 SAPD	S.aureus pdhE, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	89	76	207
4216	1	656	330	emb X58434 SAPD	S.aureus pdhE, pdhC and pdhD genes for pyruvate decarboxylase, dihydroliipoamide acetyltransferase and dihydroliipoamide dehydrogenase	98	326	327
4226	1	594	298	gb L11530	Staphylococcus aureus transfer RNA sequence with two rRNAs	97	132	297
4260	1	216	383	gb U11784	Staphylococcus aureus methicillin-resistant ATCC 33952 clone RRN40 16S-23S rRNA spacer region	83	141	168
4272	1	355	179	emb Z48003 SAUN	S.aureus gene for DNA polymerase III	100	164	177
4276	1	4	177	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	99	150	174
4277	1	1	270	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	265	270
4282	1	691	377	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	98	282	315
4291	1	379	191	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	183	189
4295	1	3	329	emb X16457 SAST	Staphylococcus aureus gene for staphylocoagulase	94	144	327
4313	1	435	280	gb L11530	Staphylococcus aureus transfer RNA sequence with two rRNAs	100	94	156
4315	1	3	185	gb J03479	S.aureus enzyme III-lac (lacP), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	100	158	183
4315	2	101	310	gb J03479	S.aureus enzyme III-lac (lacP), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	75	210
4327	1	1	294	gb U43098	Tyrosinase Tn5404 and insertion sequence IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	294	294
4360	1	603	319	gb U02910	Staphylococcus aureus ATCC 25923 16S rRNA gene, partial sequence	100	116	285
4364	1	3	146	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	95	140	146
4388	1	167	310	emb X62992 SAFN	S.aureus fnbB gene for fibronectin binding protein B	73	119	146

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
4401	1	2	313	emb X62992 SAFN	S.aureus fnbB gene for fibronectin binding protein B	97	243	312
4421	1	36	281	dbj D12572 STA2	Staphylococcus aureus rna gene for 23S ribosomal RNA	100	112	246
4426	1	3	293	emb Z18852 SACF	S.aureus gene for clumping factor	85	185	291
4428	1	493	248	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	139	246
4462	1	2	271	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	99	270	270
4466	1	1	240	emb Z18852 SACF	S.aureus gene for clumping factor	99	231	240
4469	1	1	312	gb J03479	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	99	265	312
4485	1	3	263	gb L41098	Transposon Tn5404 and insertion sequences IS1181 and IS1182 (from Staphylococcus aureus) DNA	98	259	261
4492	1	74	400	gb H8627	Staphylococcus aureus DNA gyrase B subunit (gyrB) RecF homologue (recF) and DNA gyrase A subunit (gyrA) gene, complete cds	85	104	327
4497	1	515	369	emb Z18852 SACF	S.aureus gene for clumping factor	99	213	267
4529	1	2	172	emb X64172 SARP	S.aureus rplL, orf202, rpoB(rif) and rpoC genes for ribosomal protein L7/L12, hypothetical protein ORF202, DNA-directed RNA polymerase beta & beta' chains	100	151	171
4547	1	1	300	emb X62992 SAFN	S.aureus fnbB gene for fibronectin binding protein B	100	137	300
4554	1	318	160	emb Z18852 SACF	S.aureus gene for clumping factor	84	126	159
4565	1	9	227	emb Z18852 SACF	S.aureus gene for clumping factor	84	213	219
4569	1	79	223	emb Z18852 SACF	S.aureus gene for clumping factor	98	127	144
4608	1	22	216	emb X58434 SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	92	168	195
4614	1	464	234	emb Z18852 SACF	S.aureus gene for clumping factor	86	169	231
4623	1	105	302	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	99	152	198
4632	1	18	206	gb J03479	S.aureus enzyme III-lac (lacF), enzyme II-lac (lacE), and phospho-beta-galactosidase (lacG) genes, complete cds	98	183	189
4666	1	1	222	emb Z18852 SACF	S.aureus gene for clumping factor	84	100	222
4687	1	2	166	gb J04151	S.aureus fibronectin-binding protein (fnbA) mRNA, complete cds	98	156	165

TABLE 1

S. aureus - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	match gene name	percent ident	HSP nt length	ORF nt length
4695	1	313	158	gbt114017	Staphylococcus aureus methicillin-resistance protein (mecR) gene and unknown ORF, complete cds	75	155	156
4703	1	1	153	emb158434/SAPD	S.aureus pdhB, pdhC and pdhD genes for pyruvate decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase	98	103	153

TABLE I

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	ma[cl] gene name	% sim	% ident	length (nt)
20	6	5089	4679	gi 511839	ORF1 [Staphylococcus bacteriophage phi 11]	100	100	411
149	3	2032	1577	pir 849703 8497	int gene activator RlnA - bacteriophage phi 11	100	100	456
149	5	2109	1912	gi 166161	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	100	100	198
349	2	558	409	gi 166159	Integrase (int) [Staphylococcus bacteriophage phi 11]	100	100	150
398	1	1372	707	gi 166159	Integrase (int) [Staphylococcus bacteriophage phi 11]	100	99	666
398	2	783	1001	gi 455128	exonuclease (Xis) [Staphylococcus bacteriophage phi 11]	100	100	219
502	4	1914	1764	gi 1206912	H. influenzae predicted coding region W10640 [Haemophilus influenzae]	100	71	171
849	1	2	262	gi 1373002	[polyprotein [bean common mosaic virus]	100	46	261
1349	1	277	140	gi 163359	protein synthesis initiation factor 2 (infB) [Bacillus subtilis] gi 49319	100	82	138
2880	1	21	308	gi 862933	protein kinase C inhibitor-1 [Homo sapiens]	100	98	288
3085	1	428	216	gi 1354211	PER112-like protein [Bacillus subtilis]	100	100	213
4168	2	571	398	gi 1354211	PER112-like protein [Bacillus subtilis]	100	100	174
331	1	2	247	gi 426473	nusG gene product [Staphylococcus carnosus]	98	95	246
207	2	1272	1463	gi 460259	enolase [Bacillus subtilis]	97	90	192
331	2	395	850	gi 1581638	111 protein [Staphylococcus carnosus]	97	93	456
366	1	39	215	gi 166161	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	97	95	177
680	3	718	936	gi 426473	nusG gene product [Staphylococcus carnosus]	97	97	219
357H	1	284	144	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonoma boryanum]	97	79	141
157	1	121	518	gi 1022726	unknown [Staphylococcus haemolyticus]	96	88	198
205	33	16470	16147	gi 1165302	S10 [Bacillus subtilis]	96	91	324
3919	1	48	401	gi 4971784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	96	81	334
4133	1	830	417	gi 1022726	unknown [Staphylococcus haemolyticus]	96	84	414
4168	1	708	355	gi 1354211	PER112-like protein [Bacillus subtilis]	96	95	354
4207	1	312	157	gi 602031	similar to trimethylamine DH [Mycoplama capricolus] pir 849950 849950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolus (50C3) (fragment)	96	86	156

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
4227	2	152	331	gi 871784	Clo-like ATP-dependent protease binding subunit (Bos taurus)	96	81	180
4416	1	570	286	gi 1022726	unknown (Staphylococcus haemolyticus)	96	84	285
22	1	858	430	gi 1511070	ureD (Staphylococcus xyloosus)	95	88	429
22	7	4362	4016	gi 1581787	urease gamma subunit (Staphylococcus xyloosus)	95	79	327
82	6	8794	9114	pir JG0008 JG00	ribosomal protein S7 - Bacillus stearothermophilus	95	83	321
154	9	9280	7838	gi 1354211	per112-like protein (Bacillus subtilis)	95	92	1443
186	3	2798	2055	gi 1514656	serine O-acetyltransferase (Staphylococcus xyloosus)	95	87	744
205	5	4406	4014	gi 142462	ribosomal protein S11 (Bacillus subtilis)	95	85	393
205	7	5017	4793	gi 142459	initiation factor 1 (Bacillus subtilis)	95	84	225
205	21	11365	10991	gi 1044974	ribosomal protein L14 (Bacillus subtilis)	95	93	375
259	5	7288	6644	ap P47995 YSEA	HYPOTHETICAL PROTEIN IN SECA 5'-REGION (ORF1) (FRAGMENT)	95	85	645
302	3	795	1097	gi 40186	homologous to E.coli ribosomal protein L27 (Bacillus subtilis)  i 143592 L27 ribosomal protein (Bacillus subtilis)  r C21895 C21895 ribosomal protein L27 - Bacillus subtilis p P05657 P05657_BACSU 505 RIBOSOMAL PROTEIN L27 (BL20) (BL24).  f40175 L27 gene prod	95	89	303
310	1	579	1523	gi 1177684	chorismate mutase (Staphylococcus xyloosus)	95	92	945
414	1	2	163	pir C48396 C483	ribosomal protein L34 - Bacillus stearothermophilus	95	90	162
4185	2	125	277	gi 1276841	glutamate synthase (GOGAT) (Porphyra purpurea)	95	86	153
22	2	1028	723	gi 1511069	ureF (Staphylococcus xyloosus)	94	91	306
22	5	5046	3110	gi 1410516	urease alpha subunit (Staphylococcus xyloosus)	94	85	1737
60	4	815	1372	gi 1666116	glucose kinase (Staphylococcus xyloosus)	94	87	558
205	18	10012	9516	gi 1044978	ribosomal protein S8 (Bacillus subtilis)	94	78	477
326	4	3378	2542	gi 1557492	dihydroxynaphthoic acid (DNA) synthetase (Bacillus subtilis)  i 143186 dihydroxynaphthoic acid (DNA) synthetase (Bacillus subtilis)	94	85	837
414	3	737	955	gi 1467386	thiophen and furan oxidation (Bacillus subtilis)	94	77	219
426	3	2260	1823	gi 1263908	putative (Staphylococcus epidermidis)	94	87	438
534	1	2	355	gi 1633650	enzyme II (mannitol) (Staphylococcus carnosus)	94	84	354
1017	1	2	229	gi 149435	putative (Lactococcus lactis)	94	73	228
3098	1	330	184	gi 1413952	lpa-28d gene product (Bacillus subtilis)	94	50	147

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3232	1	630	316	gi 1022725	unknown [Staphylococcus haemolyticus]	94	84	315
42	5	2089	2259	pir B48396 B483	ribosomal protein L33 - Bacillus stearothermophilus	93	81	171
101	2	1745	1383	gi 155345	arsenic efflux pump protein [Plasmid pSX267]	93	82	363
205	24	12227	11865	ap P14577 RU16	50S RIBOSOMAL PROTEIN L16	93	83	343
259	4	8291	5673	gi 499335	seach protein [Staphylococcus carnosus]	93	85	2619
275	1	2226	1114	gi 633650	enzyme II (narritol) [Staphylococcus carnosus]	93	86	1113
444	6	6207	5773	gi 1022726	unknown [Staphylococcus haemolyticus]	93	81	435
491	1	152	622	gi 46912	ribosomal protein L13 [Staphylococcus carnosus]	93	88	471
607	6	1674	2033	gi 1022726	unknown [Staphylococcus haemolyticus]	93	83	360
653	1	973	488	gi 580890	translation initiation factor IF3 (AA 1-172) [Bacillus tharothermophilus]	93	77	486
1864	1	3	194	gi 306553	ribosomal protein small subunit [Homo sapiens]	93	93	192
2997	1	28	300	gi 143390	carbamyl phosphate synthetase [Bacillus subtilis]	93	82	273
3232	2	907	596	gi 1022725	unknown [Staphylococcus haemolyticus]	93	84	312
3761	2	794	621	gi 1022725	unknown [Staphylococcus haemolyticus]	93	88	174
16	1	3	374	gi 142781	putative cytoplasmic protein; putative [Bacillus subtilis] sp P37994 UVRB_BACSU EXCINUCLEASE ABC SUBUNIT B (DNA PROTEIN) FRAGMENT1	92	83	372
31	7	5915	6124	gi 1136430	KIAA0185 protein [Homo sapiens]	92	46	210
56	19	26483	27391	gi 467401	unknown [Bacillus subtilis]	92	80	909
69	6	5882	6130	gi 530200	trophoblastin [Ovis aries]	92	53	249
145	3	2568	2018	gi 1022725	unknown [Staphylococcus haemolyticus]	92	80	531
171	3	2760	2362	gi 517475	D-amino acid transaminase [Staphylococcus haemolyticus]	92	86	399
205	12	7495	6962	gi 49189	secY gene product [Staphylococcus carnosus]	92	85	534
205	19	10812	10255	gi 1044976	ribosomal protein L5 [Bacillus subtilis]	92	82	558
219	1	710	357	gi 1303812	YgeV [Bacillus subtilis]	92	88	354
344	3	1575	1805	gi 1405474	ICspC protein [Bacillus cereus]	92	85	231
689	1	20	361	gi 413999	lpa-75d gene product [Bacillus subtilis]	92	81	342
1343	1	2	160	pir A45634 A454	ribosomal protein L19 - Bacillus stearothermophilus	92	84	159

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1958	1	524	264	gi 407908	[Efler (Staphylococcus xyloosus)]	92	80	261
3578	2	718	386	gi 1339950	[large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)]	92	78	333
3585	1	644	324	gi 1339950	[large subunit of NADH-dependent glutamate synthase (Plectonema boryanum)]	92	81	321
3640	1	4	402	gi 1022726	[unknown (Staphylococcus haemolyticus)]	92	81	399
4362	1	14	178	gi 450688	[hcdM gene of Escherichia coli product (Escherichia coli) pIR[S38437]S38437 hcdM protein - Escherichia coli pIR[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-520)]	92	76	165
4446	1	358	182	gi 1022725	[unknown (Staphylococcus haemolyticus)]	92	82	177
4549	1	462	232	gi 1022726	[unknown (Staphylococcus haemolyticus)]	92	80	231
4626	1	3	224	gi 1022725	[unknown (Staphylococcus haemolyticus)]	92	84	222
2	4	3980	4531	gi 535349	[Cdh (Bacillus subtilis)]	91	74	552
78	1	2	1126	gi 1001376	[hypothetical protein (Synecocystis sp.)]	91	78	1125
60	5	1354	1701	gi 1226043	[orf2 downstream of glucose kinase (Staphylococcus xyloosus)]	91	80	348
101	1	1989	1036	gi 150728	[arsenic efflux pump protein (Pseudomonas)]	91	80	954
187	2	412	1194	gi 142559	[ATP synthase alpha subunit (Bacillus megaterium)]	91	79	783
205	122	11579	11398	gi 40149	[S17 protein (AA 1-87) (Bacillus subtilis)]	91	83	282
206	7	8184	10262	gi 1072418	[glcA gene product (Staphylococcus carnosus)]	91	83	2079
306	2	3885	2326	gi 143012	[GMP synthetase (Bacillus subtilis)]	91	78	1560
306	3	5319	3826	gi 467399	[IMP dehydrogenase (Bacillus subtilis)]	91	79	1494
310	3	2194	3207	gi 1177685	[ccpA gene product (Staphylococcus xyloosus)]	91	81	1014
343	4	2974	3150	gi 949574	[sucrose repressor (Staphylococcus xyloosus)]	91	82	177
480	3	1606	3042	gi 433991	[ATP synthase subunit beta (Bacillus subtilis)]	91	85	1437
536	3	2026	3280	gi 143366	[adenylosuccinate lyase (PUN-8) (Bacillus subtilis) pIR[C25326]W25326 adenylosuccinate lyase (EC 4.1.2.2) - Bacillus subtilis (strain TM300)]	91	79	747
552	1	1064	615	gi 297874	[fructose-bisphosphate aldolase (Staphylococcus carnosus) pIR[M49943]M49943 fructose-bisphosphate aldolase (EC 4.1.2.13) - Staphylococcus carnosus (strain TM300)]	91	79	450
637	1	1	1536	gi 143597	[CTP synthetase (Bacillus subtilis)]	91	79	1536
859	1	21	359	gi 385178	[unknown (Bacillus subtilis)]	91	66	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1327	1	339	530	gi 496558	orfX [Bacillus subtilis]	91	71	192
2535	1	466	275	gi 511070	ureG [Staphylococcus xylosus]	91	85	132
2594	1	2	202	gi 146824	bota-cytathionase [Escherichia coli]	91	75	201
3764	1	847	425	gi 1022725	unknown [Staphylococcus haemolyticus]	91	78	423
4011	1	127	495	gi 1022726	unknown [Staphylococcus haemolyticus]	91	79	369
4227	1	1	177	gi 296464	ATPase [Lactococcus lactis]	91	66	177
42	3	815	1033	gi 520401	catalase [Haemophilus influenzae]	90	86	219
51	8	3717	4607	gi 580899	OppF gene product [Bacillus subtilis]	90	74	891
129	3	5317	4003	gi 1146206	glutamate dehydrogenase [Bacillus subtilis]	90	76	1317
164	17	16628	16933	sp P05766 RS15	30S RIBOSOMAL PROTEIN S15 (RS18)	90	74	306
171	5	2983	2819	gi 517475	D-amino acid transaminase [Staphylococcus haemolyticus]	90	78	165
205	4	4497	3550	gi 142463	RNA polymerase alpha-core-subunit [Bacillus subtilis]	90	76	948
205	6	4748	4410	gi 1044989	ribosomal protein S13 [Bacillus subtilis]	90	73	339
205	10	7165	6404	gi 49189	secY gene product [Staphylococcus carnosus]	90	81	762
205	11	6645	6472	gi 49189	secY gene product [Staphylococcus carnosus]	90	78	174
205	27	13692	13345	gi 786157	ribosomal protein S19 [Bacillus subtilis]	90	79	363
205	31	15858	15496	gi 1165303	l3 [Bacillus subtilis]	90	78	1251
260	5	7023	5773	gi 1161380	IcaA [Staphylococcus epidermidis]	90	78	570
299	6	3378	3947	gi 467440	phosphoribosylpyrophosphate synthetase [Bacillus subtilis] gi 40218 PRPP synthetase [AA 1-317] [Bacillus subtilis]	90	75	693
320	2	1025	1717	gi 312443	carbamoyl-phosphate synthase [glutamine-hydrolysing] [Bacillus aldolyticus]	90	80	189
330	4	1581	1769	gi 986963	beta-tubulin [Sporidiobolus pararoseus]	90	77	432
369	1	954	523	pir S34762 S347	L-serine dehydratase beta chain - Clostridium sp.	90	54	186
557	1	3	168	gi 1511569	M. jannaschii predicted coding region MJ1624 [Methanococcus jannaschii]	90	73	536
663	2	667	1200	gi 143786	tryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] p J17048 YMD5 tryptophan--tRNA ligase (EC 6.1.1.2) - Bacillus ubtills	90	79	261
717	1	1	261	gi 163065	hubat [Bacillus stearothermophilus]	90	81	195
745	4	1059	865	gi 1205433	H. influenzae predicted coding region H1190 [Haemophilus influenzae]	90	81	195

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1007	1	386	565	gi1143366	adenylosuccinate lyase (PUB-9) [Bacillus subtilis] pf(C29326)WBSDS	90	77	180
					adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis			
1054	1	579	331	gi1103322	OMP_729 [Escherichia coli]	90	50	249
1156	1	117	707	gi1147776	ClpP [Bacillus subtilis]	90	80	591
1180	1	408	205	gi11377831	unknown [Bacillus subtilis]	90	74	204
1253	1	1	462	gi140046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] Ir[S15936]NUSBA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	90	75	462
2951	1	3	269	gi1144816	formyltetrahydrofolate synthetase (PFHS) (tfg start codon) (EC 3.4.3) [Moorcella thermoacetica]	90	76	267
3140	1	327	166	gi11070014	protein-dependent [Bacillus subtilis]	90	52	162
4594	1	3	233	gi1871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	90	76	231
87	1	1028	1750	gi1467327	unknown [Bacillus subtilis]	89	75	723
112	1	2	505	gi1153741	ATP-binding protein [Streptococcus mutans]	89	77	504
118	1	120	398	gi11303804	YqeQ [Bacillus subtilis]	89	75	279
128	4	3545	3757	gi1460257	triase phosphate isomerase [Bacillus subtilis]	89	84	213
164	12	11667	12755	gi139954	IF2 (aa 1-741) [Bacillus stearothermophilus]	89	80	1089
205	13	7875	7405	gi1216338	OMF for L15 ribosomal protein [Bacillus subtilis]	89	76	471
205	32	16152	15823	gi11165303	L3 [Bacillus subtilis]	89	80	310
270	3	2407	2207	gi1461902[C419	arsenate reductase (EC 1.-.-.-) - Staphylococcus xyloos planaid pSX267	89	81	201
395	2	157	672	gi1520574	glutamate racemase [Staphylococcus haemolyticus]	89	80	516
494	1	3	839	gi1396259	protease [Staphylococcus epidermidis]	89	77	837
510	1	1	444	gi140046	phosphoglucose isomerase A (AA 1-449) [Bacillus stearothermophilus] Ir[S15936]NUSBA glucose-6-phosphate isomerase (EC 5.3.1.9) A - cillus stearothermophilus	89	74	444
615	1	2124	1210	gi11303812	YqeV [Bacillus subtilis]	89	74	915
841	1	18	341	gi11165303	L3 [Bacillus subtilis]	89	80	324
1111	1	352	813	gi147146	thermonuclease [Staphylococcus intermedius]	89	70	462
1875	1	2	256	gi11205108	ATP-dependent protease binding subunit [Haemophilus influenzae]	89	82	255
2963	1	11	367	gi1467458	cell division protein [Bacillus subtilis]	89	83	357

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3020	1	90	362	gi 1239988	hypothetical protein [Bacillus subtilis]	89	66	273
3565	1	2	400	gi 1236635	dihydroxy-acid dehydratase [Bacillus subtilis]	89	75	399
3586	1	105	314	gi 580832	ATP synthase subunit gamma [Bacillus subtilis]	89	82	210
3629	1	794	399	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]	89	78	396
3688	1	2	400	gi 1146206	glutamate dehydrogenase [Bacillus subtilis]	89	75	399
3699	1	794	399	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonaea boryanum]	89	75	396
4016	1	428	216	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]	89	71	213
4177	1	471	301	gi 149426	putative [Lactococcus lactis]	89	76	171
4436	1	601	302	gi 1022725	unknown [Staphylococcus haemolyticus]	89	80	300
4635	1	320	162	gi 1022725	unknown [Staphylococcus haemolyticus]	89	73	159
2	2	1330	2676	gi 520754	putative [Bacillus subtilis]	88	76	1347
42	2	468	848	ep v42321 CATA_	CATALASE (EC 1.11.1.6)	88	76	381
53	5	6389	4722	gi 474177	alpha-D-1,4-glucosidase [Staphylococcus xylosus]	88	80	1668
56	16	18018	18617	gi 467411	recombination protein [Bacillus subtilis]	88	77	600
60	3	376	843	gi 666116	glucose kinase [Staphylococcus xylosus]	88	77	468
70	2	1583	1245	gi 44095	replication initiator protein [Lactaria monocytogenes]	88	74	339
82	8	11514	12719	pir A60663 A606	translation elongation factor Tu - Bacillus subtilis	88	79	1206
103	7	4179	4391	gi 167181	serine/threonine kinase receptor [Brassica napus]	88	77	213
114	8	7732	8232	gi 1022726	unknown [Staphylococcus haemolyticus]	88	72	501
118	2	308	2011	gi 1303804	YqeQ [Bacillus subtilis]	88	77	1704
141	3	657	1136	gi 1405446	transketolase [Bacillus subtilis]	88	72	480
148	7	5871	6116	gi 1118002	dihydropterate synthase [Staphylococcus haemolyticus]	88	78	246
165	3	1428	2231	gi 40053	phenylalanyl-tRNA synthetase alpha subunit [Bacillus subtilis]	88	80	804
205	28	15027	14185	gi 1165306	tr S1730 YFBSA phenylalanine-tRNA ligase (EC 6.1.1.20) alpha ein - Bacillus subtilis			
225	1	1569	888	gi 1303840	YqfS [Bacillus subtilis]	88	78	672
235	1	2	1975	gi 452309	valyl-tRNA synthetase [Bacillus subtilis]	88	76	1974

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
339	3	2060	1586	gi 1118002	dihydropterate synthase [Staphylococcus haemolyticus]	88	73	495
443	4	4325	2928	gi 558559	pyrimidine nucleoside phosphorylase [Bacillus subtilis]	88	73	1398
532	1	3	419	gi 143797	Valyl-tRNA synthetase [Bacillus stearothermophilus] sp P1911 SV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINP--TRNA LIGASE (VALAS)	88	78	417
534	3	2504	2988	gi 153049	mammotol-specific enzyme-III [Staphylococcus carnosus] p J00088 J00088 phosphotransferase system enzyme II (EC 7.1.1.69), mammotol-specific factor III - Staphylococcus carnosus sp P17876 PTNA_STACA PTS SYSTEM, MAMMOTOL-SPECIFIC IIA COMPONENT EIIA-RTU (	88	82	465
705	2	584	399	gi 1710018	nitrite reductase (nirB) [Bacillus subtilis]	88	70	186
1000	2	1824	1309	gi 1022726	unknown [Staphylococcus haemolyticus]	88	78	516
1299	1	587	324	gi 401786	phosphomannomutase [Mycoplasma pitum]	88	55	264
1341	2	170	400	gi 39963	ribosomal protein L20 (AA 1-119) [Bacillus stearothermophilus] tr S05348 S05320 ribosomal protein L20 - Bacillus stearothermophilus	88	82	231
1386	1	41	214	gi B47154 B471	signal recognition particle 54K chain homolog f7h - Bacillus subtilis	88	71	174
1386	2	183	533	gi B47154 B471	signal recognition particle 54K chain homolog f7h - Bacillus subtilis	88	73	351
2949	1	704	399	gi 535350	CoDX [Bacillus subtilis]	88	73	306
2984	1	5	169	gi 218277	o-acetylserine(thiol) lyase [Spinacia oleracea]	88	70	165
3035	1	1	138	gi 493083	dihydroxyacetone kinase [Citrobacter freundii]	88	67	138
3089	1	3	152	gi 606055	ORF_P746 [Escherichia coli]	88	88	150
3917	1	817	410	gi 143378	pyruvate decarboxylase (E-1) beta subunit [Bacillus subtilis] gi 1377836	88	77	408
4199	1	680	342	gi 1405454	aconitase [Bacillus subtilis]	88	82	339
4201	1	734	369	gi 515938	glutamate synthase (ferredoxin) [Synecocystis sp.] p J546957 J546957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	88	84	366
4274	1	1	336	gi 515938	glutamate synthase (ferredoxin) [Synecocystis sp.] p J546957 J546957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	88	84	336
4308	1	784	399	gi 1146206	glutamate dehydrogenase [Bacillus subtilis]	88	71	396
2	5	4570	6000	gi 535350	CoDX [Bacillus subtilis]	87	70	1431
52	8	6781	6482	gi 1064791	function unknown [Bacillus subtilis]	87	66	300

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
73	3	1584	2480	gi 142992	glycerol kinase (glpk) (EC 2.7.1.30) [Bacillus subtilis] pir B45868 B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis ep P18157 GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GK).	87	72	897
98	12	8813	9100	gi 467433	unknown [Bacillus subtilis]	87	62	288
124	4	4285	2988	gi 556886	serine hydroxymethyltransferase [Bacillus subtilis] pir s49363 s49363 serine hydroxymethyltransferase - Bacillus ubtills	87	77	1278
124	6	4457	4032	gi 556883	Unknown [Bacillus subtilis]	87	66	426
148	5	3741	4559	gi 467460	unknown [Bacillus subtilis]	87	70	819
164	13	12710	13810	gi 39954	IF2 (aa 1-741) [Bacillus stearothermophilus]	87	72	1101
177	2	1104	2126	gi 467385	unknown [Bacillus subtilis]	87	78	1023
199	1	1982	1158	gi 143527	iron-sulfur protein [Bacillus subtilis]	87	77	825
199	2	4717	2933	pir A27763 A277	succinate dehydrogenase (EC 1.3.99.1) flavoprotein - Bacillus subtilis	87	80	1785
205	23	11782	11543	gi 1044972	ribosomal protein L29 [Bacillus subtilis]	87	76	240
205	25	13275	12607	gi 1165309	S3 [Bacillus subtilis]	87	75	669
222	1	2033	1107	gi 1177249	recJ3 gene product [Bacillus subtilis]	87	70	927
236	3	1635	1333	gi 1146198	ferredoxin [Bacillus subtilis]	87	80	303
246	5	2585	2392	gi 467373	ribosomal protein S16 [Bacillus subtilis]	87	77	294
260	2	4189	3422	gi 1161382	IcaC [Staphylococcus epidermidis]	87	72	768
320	3	1636	2391	gi 312443	carbamoyl-phosphate synthase (glutamine-hydrolyzing) [Bacillus aldoityticus]	87	80	696
380	4	1165	1383	gi 142570	ATP synthase c subunit [Bacillus firmus]	87	80	219
414	4	900	1073	gi 467386	thiolophen and furan oxidation [Bacillus subtilis]	87	77	174
435	2	1003	794	gi 1046166	pillin repressor [Mycoplasma genitalium]	87	69	210
448	1	1255	722	gi 405134	acetate kinase [Bacillus subtilis]	87	75	534
480	1	1	711	gi 142559	ATP synthase alpha subunit [Bacillus megaterium]	87	79	711
481	1	2	352	sp Q06797 RL1_8	[50S RIBOSOMAL PROTEIN L1 (BL1).	87	72	351
677	2	359	955	gi 460911	fructose-bisphosphate aldolase [Bacillus subtilis]	87	78	597
677	3	934	1284	gi 460911	fructose-bisphosphate aldolase [Bacillus subtilis]	87	78	351
876	1	3	452	gi 1146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	87	79	450

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match	match gene name	% sim	% ident	length (nt)
1376	1	426	214	gi 1065555	[F46H6.4 gene product [Caenorhabditis elegans]		87	75	213
2206	1	3	174	gi 215098	[exclusionase [Bacteriophage L54a]		87	72	372
2938	1	3	290	gi 508979	[GTP-binding protein [Bacillus subtilis]		87	69	288
3081	2	126	308	gi 467399	[IMP dehydrogenase [Bacillus subtilis]		87	72	183
3535	1	3	401	gi 1405454	[aconitase [Bacillus subtilis]		87	80	399
4238	1	547	275	gi 603769	[HutU protein, urocanase [Bacillus subtilis]		87	73	273
4	8	10427	8736	gi 603769	[HutU protein, urocanase [Bacillus subtilis]		86	72	1692
22	6	4190	3736	gi 410515	[urease beta subunit [Staphylococcus xylosum]		86	73	453
54	2	2480	1572	gi 289287	[UDP-glucose pyrophosphorylase [Bacillus subtilis]		86	70	909
124	3	2336	1713	gi 556887	[uracil phosphoribosyltransferase [Bacillus subtilis]		86	74	624
148	3	1349	3448	gi 467458	[uracil phosphoribosyltransferase - Bacillus subtilis]		86	75	2100
148	4	3638	3859	gi 467460	[cell division protein [Bacillus subtilis]		86	73	222
152	3	1340	2086	gi 1377835	[pyruvate decarboxylase E-1 alpha subunit [Bacillus subtilis]		86	75	747
164	18	17347	19467	gi 1184680	[polynucleotide phosphorylase [Bacillus subtilis]		86	72	2121
180	2	554	1159	gi 143467	[ribosomal protein S4 [Bacillus subtilis]		86	80	606
205	3	2966	2592	gi 142464	[ribosomal protein L17 [Bacillus subtilis]		86	77	375
205	26	13364	12990	gi 40107	[ribosomal protein L22 [Bacillus stearothermophilus]		86	75	375
246	7	3463	3140	gi 467375	[ribosomal protein L22 - Bacillus stearothermophilus]		86	70	324
299	3	1196	1540	gi 39656	[ribosomal protein S6 [Bacillus subtilis]		86	70	345
299	7	3884	4345	gi 467440	[apoVG gene product [Bacillus megaterium]		86	78	462
304	5	2170	2523	gi 666983	[phosphoribosylpyrophosphate synthetase [Bacillus subtilis]		86	78	462
310	2	1487	1678	gi 1177684	[synthetase (AA 1-317) [Bacillus subtilis]		86	65	354
337	5	2086	3405	gi 487434	[putative ATP binding subunit [Bacillus subtilis]		86	71	192
339	2	1489	1109	gi 118003	[chorismate mutase [Staphylococcus xylosum]		86	71	192
358	2	2124	3440	gi 1146219	[isocitrate dehydrogenase [Bacillus subtilis]		86	78	1320
					[dihydropyrimidin aldolase [Staphylococcus haemolyticus]		86	77	381
					[28.2% of identity to the Escherichia coli GTP-binding protein Era; putative [Bacillus subtilis]		86	73	1317

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
404	2	1015	2058	gi 1303817	YqfA [Bacillus subtilis]	86	78	1044
581	2	661	452	gi 40056	phoP gene product [Bacillus subtilis]	86	71	210
642	2	338	1075	gi 1176399	EpIF [Staphylococcus epidermidis]	86	72	738
770	1	622	347	gi 143328	phoP protein [put.]; putative [Bacillus subtilis]	86	69	276
865	1	1777	890	gi 1146247	asparaginyl-tRNA synthetase [Bacillus subtilis]	86	74	888
868	2	963	1133	gi 1002911	transmembrane protein [Saccharomyces cerevisiae]	86	69	171
904	1	1	162	gi 1303912	YohW [Bacillus subtilis]	86	72	162
989	1	35	433	gi 1303993	YokT [Bacillus subtilis]	86	76	399
1212	1	296	150	gi 414014	ipa-90d gene product [Bacillus subtilis]	86	70	147
1323	1	2	148	gi 40041	pyruvate dehydrogenase (lipoamide) [Bacillus stearothermophilus] pir S10798 DEBSP pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) pha chain - Bacillus stearothermophilus	86	75	147
3085	2	540	310	gi 1354211	PET112-like protein [Bacillus subtilis]	86	86	231
1847	1	1	228	gi 296464	ATPase [Lactococcus lactis]	86	63	228
4487	1	476	240	gi 1022726	unknown [Staphylococcus haemolyticus]	86	73	237
4583	1	372	187	gi 1022725	unknown [Staphylococcus haemolyticus]	86	79	186
25	5	4287	5039	gi 1502421	3-ketoacyl-acyl carrier protein reductase [Bacillus subtilis]	85	64	753
56	21	30637	29395	gi 1408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	85	69	1233
68	2	332	3192	gi 467376	unknown [Bacillus subtilis]	85	74	861
73	2	880	1707	gi 142992	glycerol kinase (glpk) (EC 2.7.1.30) [Bacillus subtilis] pir B45868 B45868 glycerol kinase (EC 2.7.1.30) - Bacillus subtilis ap P18157 GLPK_BACSU GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL -PHOSPHOTRANSFERASE) (GLYCEROKINASE) (GKI).	85	72	828
106	4	1505	3490	gi 143766	(thrsv) (EC 6.1.1.3) [Bacillus subtilis]	85	74	1986
128	2	1153	2202	gi 311924	glycerol dehydro-3-phosphate dehydrogenase [Clostridium pasteurianum] pir B34254 B34254 glycerol dehydro-3-phosphate dehydrogenase (EC 2.1.12) - Clostridium pasteurianum	85	75	1050
329	4	6466	5252	gi 1064807	ORNITHINE AMINOTRANSFERASE [Bacillus subtilis]	85	73	1215
338	6	3475	5673	gi 1072419	glcB gene product [Staphylococcus carnosus]	85	74	2199
189	1	2	169	gi 467385	unknown [Bacillus subtilis]	85	65	168

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
205	15	824	8106	gi 1044981	ribosomal protein S5 [Bacillus subtilis]	85	75	519
205	20	10928	10596	pir A02819 RSBS	ribosomal protein L24 - Bacillus stearothermophilus	85	72	333
220	6	6490	6101	gi 48980	secA gene product [Bacillus subtilis]	85	66	390
231	4	4877	3159	gi 1002520	MutS [Bacillus subtilis]	85	70	1719
243	9	8013	8783	gi 414011	ipa-87r gene product [Bacillus subtilis]	85	72	771
249	2	5894	3186	gi 1405454	aconitase [Bacillus subtilis]	85	73	2709
302	1	140	475	gi 40173	homolog of E. coli ribosomal protein L21 [Bacillus subtilis] ir S18439 S18439 Ribosomal protein L21 - Bacillus subtilis p P26908 RL21_BACSU 50S RIBOSOMAL PROTEIN L21 (BL20).	85	72	336
333	1	5445	2968	gi 442360	ClpC adenosine triphosphatase [Bacillus subtilis]	85	69	2478
364	6	6082	8196	gi 871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	85	68	2115
448	2	1992	1339	gi 405134	acetate kinase [Bacillus subtilis]	85	68	654
747	1	1251	853	gi 1373157	orf-X; hypothetical protein; Method: conceptual translation supplied by author [Bacillus subtilis]	85	73	399
886	2	159	467	gi 541768	hemin permease [Yersinia enterocolitica]	85	55	309
1089	1	1208	606	pir B47154 B471	signal recognition particle 54K chain homolog PfH - Bacillus subtilis	85	71	603
1163	1	816	409	gi 304155	diaminopimelate decarboxylase [Bacillus methanolicus] sp P41023 DCDA_BACMT DIAMINOPIMELATE DECARBOXYLASE (EC 4.1.1.20) DAP DECARBOXYLASE	85	62	408
1924	1	487	251	gi 215098	excisionase [Bacteriophage 154a]	85	73	237
2932	1	776	390	gi 1041099	Pyruvate Kinase [Bacillus licheniformis]	85	71	387
3030	1	3	275	gi 42370	pyruvate formate-lyase (AA 1-760) [Escherichia coli] ir S01788 S01788 formate C-acetyltransferase (EC 2.3.1.34) - cherichia coli	85	74	273
3111	1	595	299	gi 63568	limb deformity protein [Gallus gallus]	85	85	297
3778	1	630	316	gi 391840	beta-subunit of HDT [Pseudomonas fragii]	85	67	315
3835	1	1	387	gi 1204472	type I restriction enzyme ECOR124/3 I M protein [Haemophilus influenzae]	85	56	387
4042	1	3	386	gi 18178	formate acetyltransferase [Chlamydomonas reinhardtii] ir S24997 S24997 formate C-acetyltransferase (EC 2.3.1.34) - Chlamydomonas reinhardtii	85	70	384
4053	1	35	340	gi 1204472	type I restriction enzyme ECOR124/3 I M protein [Haemophilus influenzae]	85	56	306
4108	1	2	181	gi 1072418	glcA gene product [Staphylococcus carnosus]	85	61	180
4300	1	575	330	gi 151932	fructose enzyme II [Rhodobacter capsulatus]	85	59	246

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4392	1	627	1355	gi11022725	unknown [Staphylococcus haemolyticus]	85	74	273
4408	1	2	235	gi1871784	[Ctp]-like ATP-dependent protease binding subunit (Bos taurus)	85	62	234
4430	1	578	291	gi11009366	Respiratory nitrate reductase [Bacillus subtilis]	85	68	288
4555	1	2	253	gi1450688	hnm gene of Ecoprt gene product [Escherichia coli] pir[S3847/S3847] hnm protein - Escherichia coli pir[S09629/S09629] hypothetical protein A - Escherichia coli [SUB 40-520]	85	52	252
4611	1	481	242	gi11236635	dihydroxy-acid dehydratase [Bacillus subtilis]	85	65	240
4	10	10061	10591	gi146982	[foa] gene product [Staphylococcus epidermidis]	84	68	531
13	2	1348	1172	gi1162450	[ahrC] protein [Bacillus subtilis]	84	56	177
16	4	1803	4652	gi1127198	[RNA repair protein [Deinococcus radiodurans]	84	67	2850
22	3	1535	1128	gi1511069	[uraF] [Staphylococcus xylosum]	84	73	408
23	7	5055	5306	gi1603320	[Yer082p [Saccharomyces cerevisiae]	84	61	252
53	11	11597	11145	gi11303948	[YqjW [Bacillus subtilis]	84	68	453
53	12	14059	12720	gi1142613	[branched chain alpha-keto acid dehydrogenase E2 [Bacillus subtilis]	84	71	1290
70	1	1332	982	gi146647	[ORF (repE) [Staphylococcus aureus]	84	68	351
71	4	2512	4311	gi1142993	[glycerol-3-phosphate dehydrogenase (gldP) (EC 1.1.99.5) [Bacillus subtilis]	84	74	1800
98	7	4324	6096	gi1467427	[methionyl-tRNA synthetase [Bacillus subtilis]	84	66	1773
100	9	9301	8680	gi1340128	[ORF1 [Staphylococcus aureus]	84	78	822
117	3	1934	3208	gi11237019	[Srb [Bacillus subtilis]	84	68	1275
148	6	4720	5670	gi1467462	[cysteine synthetase A [Bacillus subtilis]	84	69	951
152	4	2064	2456	gi1143377	[pyruvate decarboxylase (P-1) alpha subunit [Bacillus subtilis]	84	70	393
169	7	3634	3861	gi11001342	[pir[S36718]/HSP4 pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) lpha chain - Bacillus subtilis]			
171	4	2992	2657	gi1517475	[hypothetical protein [Synecocystis sp.]	84	66	228
186	6	6941	6216	gi1467475	[D-amino acid transaminase [Staphylococcus haemolyticus]	84	71	336
205	9	6261	5692	gi1216340	[unknown [Bacillus subtilis]	84	70	726
224	2	915	1391	gi1288269	[ORF for adenylate kinase [Bacillus subtilis]	84	71	570
					[beta-fructofuranosidase [Staphylococcus xylosum]	84	70	477

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
251	1	92	388	gi 1303790	yjaZ [Bacillus subtilis]	84	65	297
282	3	1526	2836	gi 143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir 042728 042728 glutamate-1-semialdehyde 2,1-aminotransferase (EC 4.3.8) - Bacillus subtilis	84	75	1311
307	5	3138	2959	gi 1070014	protein-dependent [Bacillus subtilis]	84	62	180
320	4	2343	4229	gi 143390	[carbamyl phosphate synthetase [Bacillus subtilis]	84	70	1887
372	1	3	296	gi 1022725	unknown [Staphylococcus haemolyticus]	84	70	294
413	2	2201	1341	gi 1256146	ybbQ [Bacillus subtilis]	84	65	861
439	1	3	392	gi 1048173	osmotically inducible protein [Mycoplasma genitalium]	84	53	390
461	3	1362	2270	gi 40211	threonine synthase (thrc) (AA 1-352) [Bacillus subtilis] tr A25364 A25364 threonine synthase (EC 4.2.99.2) - Bacillus btillis	84	69	909
487	1	3	299	gi 1144531	integrin-like protein alpha Intp [Candida albicans]	84	46	297
491	2	624	905	pir S08564 R185	ribosomal protein S9 - Bacillus stearothermophilus	84	69	282
491	3	816	1033	pir S08564 R185	ribosomal protein S9 - Bacillus stearothermophilus	84	77	198
548	1	3	341	gi 411231	muacil polypeptide [Bacillus caldolyticus]	84	74	339
728	2	2701	1748	gi 912445	DNA polymerase [Bacillus caldotenax]	84	68	954
769	1	3	257	gi 1510953	cobalamin biosynthesis protein N [Methanococcus jannaschii]	84	38	255
954	1	308	156	gi 1405454	aconitase [Bacillus subtilis]	84	57	151
957	1	3	395	gi 143402	recombination protein (ctg start codon) [Bacillus subtilis] gi 130392  RecN [Bacillus subtilis]	84	68	393
975	1	3	432	gi 1885934	CipB [Synecoccus sp.]	84	70	450
1585	1	3	257	gi 1510140	ligandopeptidase P [Lactococcus lactis]	84	56	255
2954	1	3	323	gi 603769	Hutu protein, urocanase [Bacillus subtilis]	84	73	321
2996	1	650	348	gi 18178	formate acetyltransferase [Chlamydomonas reinhardtii] tr S24997 S24997 formate C-acetyltransferase (EC 2.3.1.54) - Chlamydomonas reinhardtii	84	65	303
3766	1	737	375	gi 1517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]	84	72	363
4022	1	2	169	gi 1146206	glutamate dehydrogenase [Bacillus subtilis]	84	54	168
4058	1	620	312	gi 151932	fructose enzyme II [Rhodospirillum rubrum]	84	71	309
4108	2	106	351	gi 1072418	glcA gene product [Staphylococcus carnosus]	84	77	246

TABLE 2

3 aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4183	1	3	308	gi 603769	hutU protein, urocanase [Bacillus subtilis]	84	72	306
4726	1	55	234	gi 146208	glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pifA29617 A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large hain - Escherichia coli	84	73	180
22	4	2043	1576	gi 393297	urease accessory protein [Bacillus sp.]	83	64	468
53	13	14722	13745	gi 142612	branched chain alpha-keto acid dehydrogenase E1-beta [Bacillus subtilis]	83	68	978
57	16	13357	12872	gi 143132	lactate dehydrogenase (AC 1.1.1.27) [Bacillus caldolyticus] pifB29704 B29704 L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus caldolyticus	83	66	486
66	3	3119	2274	gi 1303694	yqkM [Bacillus subtilis]	83	63	866
66	5	6118	4643	gi 1212730	yqkK [Bacillus subtilis]	83	68	1476
70	3	1864	1523	gi 144095	replication initiator protein [Bacteria monocytogenes]	83	73	342
90	1	377	1429	gi 155571	alcohol dehydrogenase I (adhA) (EC 1.1.1.1) [Zymomonas mobilis] pifA3260 A3260 alcohol dehydrogenase (EC 1.1.1.1) I - Zymomonas mobilis	83	70	1053
95	2	708	2162	gi 506381	phospho-beta-glucosylase [Bacillus subtilis]	83	70	1455
137	1	68	694	gi 467391	initiation protein of replication [Bacillus subtilis]	83	77	677
140	4	3209	2742	gi 634107	kdpB [Escherichia coli]	83	65	466
142	3	3468	2989	gi 1212776	lumazine synthase (b-subunit) [Bacillus amyloquelaciensis]	83	69	480
161	12	5749	4696	gi 903307	ORF75 [Bacillus subtilis]	83	64	948
164	9	9480	11070	gi 49316	ORF2 gene product [Bacillus subtilis]	83	66	1191
164	14	14148	14546	gi 580902	ORF6 gene product [Bacillus subtilis]	83	60	399
170	2	3144	2467	gi 520846	orf4 [Bacillus subtilis]	83	64	678
186	2	2029	1370	gi 289284	cysteinyI-tRNA synthetase [Bacillus subtilis]	83	72	660
205	14	7822	7607	gi 216337	ORP for L30 ribosomal protein [Bacillus subtilis]	83	74	216
217	6	3683	4540	gi 1510488	imidazoleglycerol-phosphate synthase (cyclase) [Methanococcus jannaschii]	83	60	858
301	1	985	638	gi 467419	unknown [Bacillus subtilis]	83	65	348
302	4	1421	2743	gi 508979	GTP-binding protein [Bacillus subtilis]	83	68	1323
321	4	3933	3571	gi 39843	fumarate (citG) (aa 1-462) [Bacillus subtilis]	83	68	363
367	1	2	352	gi 1039479	ORF1 [Lactococcus lactis]	83	54	351

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
387	1	3	662	gi 806281	DNA polymerase I [Bacillus stearothermophilus]	83	70	660
527	2	916	1566	gi 396259	protease [Staphylococcus epidermidis]	83	67	651
533	1	355	179	gi 142455	alanine dehydrogenase (EC 1.4.1.1) [Bacillus stearothermophilus] pir B14261 B14261 alanine dehydrogenase (EC 1.4.1.1) - Bacillus stearothermophilus	83	66	177
536	4	1617	1438	gi 143366	adenylosuccinate lyase (PDB-8) [Bacillus subtilis]   r[C29336]WZSDS adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	83	67	180
652	1	2	859	gi 520753	DNA topoisomerase I [Bacillus subtilis]	83	72	858
774	2	200	361	gi 1522665	M. jannaschii predicted coding region MJCU28 [Methanococcus jannaschii]	83	58	162
897	1	120	296	gi 1064807	ORTHONINE AMINOTRANSFERASE [Bacillus subtilis]	83	76	177
1213	1	3	491	gi 289288	lexA [Bacillus subtilis]	83	67	489
2529	1	296	150	gi 143786	cryptophanyl-tRNA synthetase (EC 6.1.1.2) [Bacillus subtilis] pir J70481 J70481 tRNA synthetase (EC 6.1.1.2) - Bacillus subtilis	83	69	147
2973	1	649	326	gi 1109687	ProZ [Bacillus subtilis]	83	58	324
1009	1	728	106	gi 148232	ORF_0294 [Bacillus thuringiensis]	83	65	163
3035	2	45	305	gi 950062	hypothetical yeast protein 1 [Hypoclema capricolum] pir S45578 S45578 hypothetical protein - Mycoplasma capricolum 50C1 (fragment)	83	59	261
1906	1	67	302	gi 335197	Chloroform reductase [Bacterium acidaminophilum]	83	61	243
445R	1	540	271	gi 397526	clumping factor [Staphylococcus aureus]	83	78	270
4570	1	444	223	gi 1022726	unknown [Staphylococcus haemolyticus]	83	74	222
4654	1	97	261	gi 1072419	glcB gene product [Staphylococcus carnosus]	83	79	165
16	2	295	1191	gi 153854	juv402 protein [Streptococcus pneumoniae]	82	67	897
16	3	1193	1798	gi 153854	juv402 protein [Streptococcus pneumoniae]	82	70	606
38	12	9644	8724	gi 1204400	N-acetylneuraminidase lyase [Haemophilus influenzae]	82	58	921
42	4	988	2019	gi 841192	catalase [Bacteroides fragilis]	82	70	1032
51	6	2590	3489	gi 143607	sporulation protein [Bacillus subtilis]	82	69	900
56	11	12270	13925	gi 39431	oligo-1,6-glucosidase [Bacillus cereus]	82	60	1656
56	15	17673	18014	gi 467410	unknown [Bacillus subtilis]	82	66	342
61	2	881	3313	gi 143148	transfer RNA-Leu synthetase [Bacillus subtilis]	82	70	2433

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
82	7	9162	11318	gi 48240	elongation factor G (AA 1-691) [Thermus aquaticus thermophilus] pIR[S1528]EFMD elongation factor G - Thermus aquaticus pIR[S1551]EPG_THETH ELONGATION FACTOR G (EP-Q).	82	64	2157
85	2	5470	3260	gi 143369	phosphoribosylformyl glycylamide synthetase II (PUR-Q) [Bacillus subtilis]	82	66	2211
102	6	3662	5380	gi 1256635	dihydroxy-acid dehydratase [Bacillus subtilis]	82	65	1719
117	4	3242	3493	pir[A47154]A471	orf1 9' of Ffh - Bacillus subtilis	82	53	252
128	6	4177	5933	gi 460258	phosphoglycerate mutase [Bacillus subtilis]	82	66	1557
129	2	1229	2182	gi 403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pIR[S1725][S1725] glycerophosphoryl diester phosphodiesterase - acillus subtilis	82	62	954
170	1	2	1441	gi 1377831	unknown [Bacillus subtilis]	82	67	1440
177	1	3	1094	gi 467386	thiophen and furan oxidation [Bacillus subtilis]	82	65	1092
184	4	3572	4039	gi 153566	ORP (19K protein) [Enterococcus faecalis]	82	59	468
189	8	4455	4225	gi 1001878	CapB protein [Listeria monocytogenes]	82	73	231
206	19	21366	20707	gi 473916	lipopeptide antibiotics Iturin A [Bacillus subtilis] epIR[S1914]LPI4_BACSU LITROPEPTIDE ANTIBIOTICS ITURIN A AND SURFACTIN BIOSYNTHESIS PROTEIN.	82	50	660
221	2	805	1722	gi 517205	167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	82	63	918
223	4	3866	3651	gi 439619	[Salmonella typhimurium] IS200 insertion sequence from SARA17, artial.], gene product [Salmonella typhimurium]	82	69	216
260	3	5207	4296	gi 1161381	Ic6B [Staphylococcus epidermidis]	82	61	912
315	3	4864	2855	gi 143397	guinol oxidase [Bacillus subtilis]	82	67	2010
321	10	8520	7945	gi 142981	ORF5; This ORF includes a region (aa23-103) containing a potential non-sulphur centre homologous to a region of Rhodospirillum rubrum and Chromatium vinosum; putative [Bacillus stearothermophilus] pIR[PQ0299]PQ0299 hypothetical protein 5 (gldA 3' region) -	82	62	576
331	3	1055	1342	gi 436574	ribosomal protein L1 [Bacillus subtilis]	82	71	288
370	2	262	618	gi 1303793	YqeL [Bacillus subtilis]	82	59	357
404	4	3053	4024	gi 1303821	YqeE [Bacillus subtilis]	82	68	972
405	4	4440	3073	gi 1303913	YqhX [Bacillus subtilis]	82	67	1368
436	3	4096	2864	gi 149521	cryptophan synthase beta subunit [Bacillus subtilis] pIR[S15129]S15129 cryptophan synthase (EC 4.2.1.20) beta chain - actococcus lactis subsp. lactis	82	67	1233

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
441	4	3394	2573	gi142952	glyceraldh/de-3-phosphate dehydrogenase [Bacillus teurothermophilus]	82	67	822
444	12	10415	11227	gi1204354	spore germination and vegetative growth protein [Haemophilus influenzae]	82	67	813
446	1	3	191	gi143387	aspartate transcarbamylase [Bacillus subtilis]	82	66	189
462	3	1007	1210	gi142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] pir1437192 A37192 uvrB protein - Bacillus subtilis ap1437192 UVRB_BACSU ENACNUCLEASE ABC SUBUNIT C	82	64	204
537	1	1560	784	gi1853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]	82	61	777
680	2	407	700	gi1426472	secE gene product [Staphylococcus carnosus]	82	69	294
724	2	565	386	gi143373	phosphoribosyl aminimidazole carboxy forsy/ ormyltransferase/inosine monophosphate cyclodiphosphate (Pur-H37) Bacillus subtilis	82	68	180
763	1	422	213	gi1467458	cell division protein [Bacillus subtilis]	82	35	210
818	1	564	283	gi11064787	function unknown [Bacillus subtilis]	82	69	282
858	1	175	1176	gi143043	uroporphyrinogen decarboxylase [Bacillus subtilis] pir147045 B47045 uroporphyrinogen decarboxylase [EC 4.1.1.37] - acillus subtilis	82	71	1002
885	1	3	599	gi11027507	ATP binding protein [Borrelia burgdorferi]	82	72	597
939	1	10	399	gi143795	transfer RNA-Tyr synthetase [Bacillus subtilis]	82	60	390
961	1	1	306	gi1577647	gamma-hemolysin [Staphylococcus aureus]	82	69	306
1192	1	307	155	gi146974	NH3-dependent NAD synthetase [Escherichia coli]	82	71	153
1317	1	49	375	gi1407908	glucanase [Staphylococcus xylosus]	82	72	327
1341	1	1	150	gi139962	ribosomal protein L35 (AA 1-66) [Bacillus teurothermophilus]	82	68	150
2990	2	367	349	gi1534855	ATPase subunit epsilon [Bacillus teurothermophilus] ap142009 ATPE_BACST ATP SYNTHASE EPSILON CHAIN [EC 3.6.1.34]	82	47	219
3024	1	45	224	gi1467402	unknown [Bacillus subtilis]	82	64	180
3045	1	276	139	gi1467335	ribosomal protein L9 [Bacillus subtilis]	82	60	138
3045	2	558	400	gi1467335	ribosomal protein L9 [Bacillus subtilis]	82	82	159
3091	1	474	238	gi1499335	secA protein [Staphylococcus carnosus]	82	78	237
3107	1	416	210	gi1546918	orfV 3' of conk [Bacillus subtilis, E26, Peptide Partial, 140 aa] pir1431612 S43612 hypothetical protein Y - Bacillus subtilis ap140336 THAD_BACSU HYPOTHETICAL PROTEIN IN CONK 3'REGION (ORFV FRAGMENT)	82	64	207

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4332	1	2	319	gi 42086	nitrate reductase alpha subunit [Escherichia coli] p[09152]NARO_ECOLI RESPIRATORY NITRATE REDUCTASE 1 ALPHA CHAIN (EC 7.99.4). (SUB 2-1247)	82	75	318
23	3	3275	2574	gi 1199573	spae [Sphingomonas sp.]	81	64	702
42	1	638	321	gi 466778	lysine specific permease [Escherichia coli]	81	59	318
48	5	4051	4350	gi 1045937	M. genitalium predicted coding region MG246 [Mycoplasma genitalium]	81	62	300
51	4	1578	2579	gi S16649/S166	delac protein - Bacillus subtilis	81	55	1002
53	2	354	1494	gi 1303961	lyqJ [Bacillus subtilis]	81	67	1131
53	8	9419	7871	gi 1146930	16-phosphogluconate dehydrogenase [Escherichia coli]	81	66	1449
54	9	10757	10119	gi 143016	permease [Bacillus subtilis]	81	65	639
54	10	13360	11786	gi 143015	glucuronate kinase [Bacillus subtilis]	81	64	1575
57	17	13983	13366	gi A23805/A258	1-L-lactate dehydrogenase (EC 1.1.1.27) - Bacillus subtilis	81	74	618
81	2	2708	2217	gi 1223302	NiFeU-related protein [Haemophilus influenzae]	81	54	492
86	1	745	374	gi 414017	lpa-93d gene product [Bacillus subtilis]	81	70	372
103	6	6438	4861	gi 971342	nitrate reductase beta subunit [Bacillus subtilis] sp[P42176]NARR_BACSU NITRATE REDUCTASE BETA CHAIN (EC 1.7.99.4)	81	64	1578
120	15	10845	12338	gi 1524392	GbsA [Bacillus subtilis]	81	67	1494
128	5	3676	4413	gi 143319	triose phosphate isomerase [Bacillus megaterium]	81	64	718
131	9	10308	9280	gi 299163	alanine dehydrogenase [Bacillus subtilis]	81	68	1039
143	6	6088	5471	gi 439619	[Salmonella typhimurium] IS700 insertion sequence from SRA17, attB1.1, gene product [Salmonella typhimurium]	81	61	618
169	1	43	825	gi 897795	30S ribosomal protein [Pedococcus acidilactici] sp[P49668]RS2_PEDAC 30S RIBOSOMAL PROTEIN S2	81	65	783
210	1	450	226	gi 1125826	short region of weak similarity to tyrosine-protein kinase receptors in a fibronectin type III-like domain [Caenorhabditis elegans]	81	54	225
213	5	2000	2677	gi 467404	unknown [Bacillus subtilis]	81	63	678
241	2	3081	2149	gi 16510	succinate-CoA ligase (GDP-forming) [Arabidopsis thaliana] tr[S10579]S10579 succinate-CoA ligase (GDP-forming) (EC 6.2.1.4) pha chain - Arabidopsis thaliana (fragment)	81	69	933
256	1	1	981	gi S09411/S094	spoIIIE protein - Bacillus subtilis	81	65	981
259	3	3752	2691	sp P28367 NP2_8	PROBABLE PEPTIDE CHAIN RELEASE FACTOR 2 (RF-2) (FRAGMENT)	81	65	1062

TABLE 2

3. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
275	2	1728	3581	gi 726480	L-glutamine-D-fructose-6-phosphate amidotransferase (Bacillus subtilis)	81	68	1854
285	1	1466	735	gi 1204844	H. influenzae Predicted coding region H10594 (Haemophilus influenzae)	81	63	732
296	1	99	1406	gi 467328	adenylosuccinate synthetase (Bacillus subtilis)	81	67	1308
302	9	5590	5889	gi 147485	queA (Escherichia coli)	81	64	300
317	2	1137	1376	gi 154961	resolvase (Transposon Tn917)	81	57	240
343	2	1034	1342	gi 405955	yeoD (Escherichia coli)	81	60	309
360	2	1404	2471	gi 1204570	aspartyl-tRNA synthetase (Haemophilus influenzae)	81	67	1068
364	5	6251	5706	gi 1204652	methylated-DNA--protein-cysteine methyltransferase (Haemophilus influenzae)	81	63	546
372	2	1707	1135	gi 467416	unknown (Bacillus subtilis)	81	65	573
392	1	43	603	pir S09411 S094	spoIIIE protein - Bacillus subtilis	81	65	561
404	9	5252	6154	gi 606745	Bex (Bacillus subtilis)	81	65	903
426	2	1727	1119	gi 39453	Manganese superoxide dismutase (Bacillus caldotenax) tr S22053 S22053 superoxide dismutase (EC 1.15.1.1) (Mn) - Bacillus licheniformis	81	66	609
480	7	4653	5089	pi C370A C370	hypothetical protein II (cpgII 3' region) - Salmonella typhimurium (fragment)	81	57	217
625	3	1105	2070	gi 1262360	protein kinase PknB (Mycobacterium leprae)	81	56	966
754	2	504	1064	gi 1303902	YqhW (Bacillus subtilis)	81	71	561
842	1	86	430	gi 1403446	transketolase (Bacillus subtilis)	81	68	345
953	1	798	400	gi 1205429	dipeptide transport ATP-binding protein (Haemophilus influenzae)	81	57	399
961	2	252	401	gi 487686	synergohymenotropic toxin (Staphylococcus intermedius) pir S46946 S46944 synergohymenotropic toxin - Staphylococcus intermedius	81	72	150
1035	1	1	189	gi 1046138	M. genitalium predicted coding region MG423 (Mycoplasma genitalium)	81	43	189
1280	1	670	449	gi 559164	halicase (Autographa californica nuclear polyhedrosis virus) ap P24307 P13.MPVAC HELICASE	81	43	222
3371	1	68	241	gi 1322245	mevalonate pyrophosphate decarboxylase (Rattus norvegicus)	81	62	174
3715	1	475	239	gi 537137	ORP_1388 (Escherichia coli)	81	58	237
3908	1	2	325	gi 439619	(Salmonella typhimurium) IS200 insertion sequence from SNA17, attL1, gene product (Salmonella typhimurium)	81	68	324
3940	1	3	401	gi 1296464	ATPase (Lactococcus lactis)	81	69	399

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match region	match gene name	% sim	% ident	length (nt)
3954	1	1	318	gi1224069	amidase [Moraxella catarrhalis]	81	68	318
4049	1	337	170	gi1603768	Hut1 protein, imidazole-5-propionate hydrolase [Bacillus subtilis] gi1603768 Hut1 protein, imidazole-5-propionate hydrolase [Bacillus subtilis]	91	68	168
4209	1	1	324	gi1403373	glycerophosphoryl diester phosphodiesterase [Bacillus subtilis] pir[S16013]S16013 pir[S16013]S16013	81	58	324
4371	1	627	322	gi1216677	indolepyruvate decarboxylase [Enterobacter cloacae] pir[S16013]S16013 indolepyruvate decarboxylase (EC 4.1.1.-) - nterobacter cloacae	81	72	306
4387	1	19	228	gi1460689	TVO (Thermosactinomyces vulgaris)	81	59	210
4191	1	581	306	gi11524193	unknown [Mycobacterium tuberculosis]	81	67	276
4425	1	3	341	gi1143015	glucanase kinase [Bacillus subtilis]	81	66	339
9	1	1593	847	gi11064786	function unknown [Bacillus subtilis]	80	62	747
17	1	544	311	gi1559164	helicase [Autographa californica nuclear polyhedrosis virus] ap[P24307]V143.NPVAC HBLCAS2	80	40	234
45	2	1159	2448	gi1109684	proV [Bacillus subtilis]	80	63	1290
45	5	4032	4733	gi1109687	proZ [Bacillus subtilis]	80	55	702
54	8	10266	9502	gi1563952	glucanase perasease [Bacillus licheniformis]	80	52	765
62	12	8852	7545	gi1854655	Na/H antiporter system [Bacillus alcalophilus]	80	62	1308
62	14	8087	8683	gi1559713	ORF (Homo sapiens)	80	68	597
67	16	13781	14122	gi1305002	ORF_E356 [Escherichia coli]	80	65	342
70	13	11495	10296	gi1303995	VqM [Bacillus subtilis]	80	64	1200
98	9	6136	7130	gi1467428	unknown [Bacillus subtilis]	80	68	795
98	10	7294	7833	gi1467430	unknown [Bacillus subtilis]	80	64	540
98	11	7820	8737	gi1467431	high level kasamycin resistance [Bacillus subtilis]	80	61	918
109	16	14154	14813	gi1580875	lipa-57d gene product [Bacillus subtilis]	80	63	660
112	15	14294	16636	gi11072361	pyruvate-formate-lyase [Clostridium pasteurianum]	80	65	2343
139	1	1448	726	gi1506699	capC [Staphylococcus aureus]	80	58	723
139	2	2179	1448	gi1506698	capB [Staphylococcus aureus]	80	59	732
174	4	3271	2870	gi1146242	aspartate 1-decarboxylase [Bacillus subtilis]	80	61	402

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
177	3	2102	2842	gi1467385	unknown [Bacillus subtilis]	80	70	741
184	6	6124	5912	gi1161953	85-kDa surface antigen [Trypanosoma cruzi]	80	46	213
186	4	5368	3875	gi1289282	glutamyl-tRNA synthetase [Bacillus subtilis]	80	65	1494
205	30	15796	15140	gi140103	ribosomal protein L4 [Bacillus stearothermophilus]	80	66	657
207	1	140	1315	gi1460259	enolase [Bacillus subtilis]	80	67	1176
211	3	1078	1590	gi1410131	ORFX7 [Bacillus subtilis]	80	61	513
235	2	1962	2255	gi1143797	valyl-tRNA synthetase [Bacillus stearothermophilus] sp P1931 SVV_BACST VALYL-TRNA SYNTHETASE (EC 6.1.1.9) VALINE--TRNA LIGASE (VALRS)	80	55	294
239	1	1	1263	gi1343000	proton glutamate symport protein [Bacillus stearothermophilus] pir S26247 S26247 glutamate/aspartate transport protein - Bacillus stearothermophilus	80	59	1263
272	5	2724	2461	gi1709993	hypothetical protein [Bacillus subtilis]	80	54	264
301	3	1446	1111	gi1467418	unknown [Bacillus subtilis]	80	58	336
310	4	5697	4501	gi1177486	lacC gene product [Staphylococcus xylosum]	80	67	1197
310	6	5358	7006	gi1348053	acetyl-CoA synthetase [Bacillus subtilis]	80	67	1749
310	7	7410	9113	gi1103865	formyl-tetrahydrofolate synthetase [Streptococcus mutans]	80	67	1704
325	3	1114	1389	gi1310325	outer capsid protein [Rotavirus sp.]	80	40	276
337	1	1268	636	gi1537049	ORF_0470 [Escherichia coli]	80	55	633
374	2	939	1228	gi13405448	YnfF [Bacillus subtilis]	80	70	300
375	5	3062	3331	gi1467448	unknown [Bacillus subtilis]	80	68	270
388	1	267	587	gi11064791	function unknown [Bacillus subtilis]	80	65	321
394	1	9	659	gi1304976	matches PS00017: ATP GTP_A and PS00301: EFACITOR GTP: similar to longation factor G, TetM/TetO tetracycline-resistance proteins [Escherichia coli]	80	65	651
456	1	625	1263	gi11146183	putative [Bacillus subtilis]	80	65	639
475	1	1	654	gi1288269	beta-fructofuranosidase [Staphylococcus xylosum]	80	66	654
544	2	1449	2240	gi1529754	apeC [Streptococcus pyogenes]	80	50	792
622	4	1623	1821	gi11483545	unknown [Mycobacterium tuberculosis]	80	65	249
719	1	1	1257	gi11064791	function unknown [Bacillus subtilis]	80	68	1257
739	1	107	838	gi1666983	putative ATP binding subunit [Bacillus subtilis]	80	61	732

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession (/)	match gene name	% sim	% ident	length (nt)
745	2	581	414	gi 1511600	coenzyme PQQ synthesis protein III [Methanococcus jannaschii]	80	61	168
822	1	17	679	gi 410141	ORF17 [Bacillus subtilis]	80	68	663
827	2	991	836	gi 1205301	leukotoxin secretion ATP-binding protein [Haemophilus influenzae]	80	54	156
1044	1	3	149	gi 160632	vp2 [Marburg virus]	80	55	147
1220	2	971	413	pir A61072 EPSG	gallicidamin precursor - Staphylococcus gallinarum	80	74	159
2519	1	75	275	gi 147536	cdp [Escherichia coli]	80	45	201
2847	1	503	279	gi 1184680	polynucleotide phosphorylase [Bacillus subtilis]	80	62	225
3120	1	2	226	gi 517205	[67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]	80	65	225
3191	1	294	148	gi 151359	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas aeruginosa] pir A44756 A44756 hydrosymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	80	59	147
3560	2	285	434	gi 217130	photosystem I core protein B [Synecoccus vulcanus]	80	70	150
3655	1	47	346	gi 415855	deoxyribose aldolase [Mycoplasma hominis]	80	56	300
3658	2	324	584	gi 551531	2-nitropropane dioxygenase [Mycobacterium tuberculosis]	80	54	261
3769	1	798	400	gi 133950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	80	68	399
3781	1	692	348	gi 166412	NADH-glutamate synthase [Medicago sativa]	80	62	345
3988	1	48	287	gi 1204696	fructose-6-phosphate 1,6-bisphosphate carboxylase [Haemophilus influenzae]	80	69	240
4030	1	571	287	gi 1009366	respiratory nitrate reductase [Bacillus subtilis]	80	60	285
4092	1	547	275	gi 1370207	orf6 [Lactobacillus sake]	80	69	273
4103	1	680	342	gi 39956	lglc [Bacillus subtilis]	80	65	339
4231	1	692	348	gi 289287	UDP-glucose pyrophosphorylase [Bacillus subtilis]	80	65	345
4265	1	595	299	gi 603768	HutI protein, imidazolone-5-propiolate hydrolase [Bacillus subtilis] gi 603768 HutI protein, imidazolone-5-propiolate hydrolase [Bacillus subtilis]	80	63	297
4504	1	498	250	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	80	68	249
2	6	5998	6798	gi 535351	CodY [Bacillus subtilis]	79	63	801
4	7	8295	7037	gi 603768	HutI protein, imidazolone-5-propiolate hydrolase [Bacillus subtilis] gi 603768 HutI protein, imidazolone-5-propiolate hydrolase [Bacillus subtilis]	79	64	1245
25	6	5273	5515	pir A36728 A367	acyl carrier protein - Rhizobium meliloti	79	65	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
59	2	1173	1424	gi 147923	threonine dehydratase 2 [EC 4.2.1.16] [Escherichia coli]	79	75	252
60	1	1	204	gi 666115	orf1 upstream of glucose kinase [Staphylococcus xyloso] pir[S52351]S52351 hypothetical protein 1 - Staphylococcus xyloso	79	60	204
61	1	3002	1590	gi 466982	[ppp1: B1996_C2_189 [Mycobacterium leprae]	79	64	1413
65	7	7023	6505	gi 143364	[phosphoribosyl aminimidazole carboxylase 1 (PUR-E) [Bacillus subtilis]	79	60	519
89	6	5660	4554	gi 144906	product homologous to E.coli thiorodoxin reductase; J.Biol.Chem. 1988) 263:9015-9019, and to F52a protein of alky hydroperoxide oxidase from S.typhimurium; J.Biol.Chem. (1990) 265:10515-10540; pen reading frame A [Clostridium pasteurianum]	79	35	1107
102	11	7489	8571	gi 143093	ketol-acid reductoisomerase [Bacillus subtilis] sp P37253 ILVC_BACSU_VETOU-ACID REDUCTOISOMERASE [EC 1.1.1.86] ACETONIDROXY-ACID ISOMEROREDUCTASE] (ALPHA-KETO-BETA-HYDROXYLACIL REDUCTOISOMERASE)	79	64	1083
102	14	11190	12563	gi 149428	putative [Lactococcus lactis]	79	65	1374
127	9	7792	9372	gi 458668	[PrfC/R3] [Dichelobacter nodosus]	79	68	1581
139	3	2540	3993	gi 506697	[CapA [Staphylococcus aureus]	79	55	558
144	2	1644	1156	gi 1498296	[peptide methionine sulfoxide reductase [Streptococcus pneumoniae]	79	47	489
148	2	529	1098	gi 467457	[hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis] gi 467457 hypoxanthine-guanine phosphoribosyltransferase [Bacillus subtilis]	79	59	570
150	1	965	591	gi 755602	unknown [Bacillus subtilis]	79	61	375
176	1	1019	587	gi 297874	[fructose-bisphosphate aldolase [Staphylococcus carnosus] pir A49943 A49943 fructose-bisphosphate aldolase [EC 4.1.2.13] - taphylococcus carnosus (strain TM300)	79	65	453
186	7	7584	6874	gi 1314298	[ORF5: putative Sma protein; similar to Sma proteins from Haemophilus influenzae and Escherichia coli [Listeria monocytogenes]	79	64	711
205	16	8887	8498	gi 1044980	[ribosomal protein L18 [Bacillus subtilis]	79	70	390
211	1	1	519	gi 3303994	[YqkM [Bacillus subtilis]	79	62	519
223	2	4183	2801	gi 488430	[alcohol dehydrogenase 2 [Entamoeba histolytica]	79	60	1383
243	8	8915	7896	gi 580883	[lpa-88d gene product [Bacillus subtilis]	79	60	1020
279	4	3721	4329	gi 413930	[lpa-6d gene product [Bacillus subtilis]	79	59	609
300	1	11	1393	gi 403372	[glycerol 3-phosphate permease [Bacillus subtilis]	79	62	1383
307	3	2930	1935	gi 1950062	[hypothetical yeast protein 1 [Mycoplasma capricolum] pir S48578 S48578 hypothetical protein - Mycoplasma capricolum SOC3] (fragment)	79	60	996

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
352	6	110106	8886	gi1216854	1947X [Pseudomonas chlororaphis]	79	59	1221
412	1	1153	578	gi1143177	putative [Bacillus subtilis]	79	51	576
481	3	621	1124	gi1786163	Ribosomal Protein L10 [Bacillus subtilis]	79	66	504
516	1	702	352	gi1805090	NisA [Lactococcus lactis]	79	48	351
525	2	2457	1426	gi1143371	phosphoribosyl aminimidazole synthetase (PUR-M) [Bacillus subtilis] pir1429326 [AJ55CU phosphoribosylformylglycinamide cyclo-ligase BC 6.3.3.1] - Bacillus subtilis	79	61	1032
538	4	3448	2825	gi11370207	lorf6 [Lactobacillus sake]	79	67	624
570	1	2	421	gi1476160	arginine permease substrate-binding subunit [Listeria monocytogenes]	79	61	420
645	8	2663	3241	gi1153898	transport protein [Salmonella typhimurium]	79	62	579
683	1	75	374	gi11064795	function unknown [Bacillus subtilis]	79	62	300
816	3	4700	3987	gi11407784	lorf-1: novel antigen [Staphylococcus aureus]	79	62	714
2929	1	3	401	gi11524397	glycine betaine transporter Opd [Bacillus subtilis]	79	61	399
2937	1	357	202	gi155291515529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	79	58	156
2940	1	768	385	gi1149429	putative [Lactococcus lactis]	79	72	384
2946	1	570	286	gi1143287	2-oxoglutarate dehydrogenase (odhA, EC 1.2.4.2) [Bacillus subtilis]	79	61	285
2999	1	3	212	gi1710020	nitrite reductase (nirB) [Bacillus subtilis]	79	59	210
3022	1	514	332	gi1450686	3-phosphoglycerate kinase [Thermotoga maritima]	79	61	183
3064	1	3	314	gi11204436	pyruvate formate-lyase [Haemophilus influenzae]	79	60	312
3083	1	2	220	gi1149662	hlypD gene product [Clostridium perfringens]	79	56	219
3126	1	701	411	gi11339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	79	55	291
3181	1	607	326	gi11339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	79	59	282
3345	1	3	476	gi1871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	79	63	474
3718	1	536	270	gi11368891368	leus protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)	79	71	267
3724	2	159	401	gi11009366	respiratory nitrate reductase [Bacillus subtilis]	79	64	243
3836	1	608	312	gi11524193	unknown [Mycobacterium tuberculosis]	79	65	297
3941	1	2	334	gi1415855	deoxyribose aldolase [Mycoplasma hominis]	79	54	333
4113	3	3	341	gi1143015	gluconate kinase [Bacillus subtilis]	79	63	339

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4501	1	406	209	gi 1022726	unknown [Staphylococcus haemolyticus]	79	66	198
4612	1	2	238	gi 406089	TVG [Thermactinomyces volgaris]	79	58	237
2	1	2	1213	gi 520753	DNA topoisomerase I [Bacillus subtilis]	78	64	1212
8	2	2266	1220	gi 216151	DNA polymerase (gene 1; ttg start codon) [Bacteriophage SP02] gi 579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] p1c[A21498]DUBP52 DNA-directed DNA polymerase (EC 2.7.7) - phage P02	78	72	1047
9	2	1340	1089	gi 1044787	function unknown [Bacillus subtilis]	78	57	252
32	8	6803	7702	gi 146974	NH3-dependent NAD synthetase [Escherichia coli]	78	63	900
36	4	2941	3138	gi 290503	glutamate permease [Escherichia coli]	78	53	198
53	15	117684	16221	gi 1303941	Vqiv [Bacillus subtilis]	78	58	1464
57	14	10520	12067	gi 1072418	glcA gene product [Staphylococcus carnosus]	78	65	1548
66	7	6798	5812	gi 1212729	VqN2 [Bacillus subtilis]	78	67	987
67	4	4029	4376	gi 466612	nika [Escherichia coli]	78	71	368
91	9	10058	10942	gi 467380	stage 0 sporulation [Bacillus subtilis]	78	50	885
102	12	8774	10130	gi 149426	putative [Lactococcus lactis]	78	61	1557
112	6	3540	4463	gi 854234	cymC gene product [Klebsiella oxytoca]	78	56	924
124	2	1888	1061	gi 405622	unknown [Bacillus subtilis]	78	60	828
130	3	1805	2260	gi 1256636	putative [Bacillus subtilis]	78	71	456
133	3	751	377	gi 168060	lamB [Escherichia nidulans]	78	59	375
166	4	7125	6163	gi 451216	mannosephosphate isomerase [Streptococcus mutans]	78	63	963
186	1	1586	795	gi 289284	cysteinyI-tRNA synthetase [Bacillus subtilis]	78	63	792
195	4	2749	2315	gi 1353874	unknown [Rhodobacter capsulatus]	78	58	435
199	3	4279	1623	gi 143525	succinate dehydrogenase cytochrome b-558 subunit [Bacillus subtilis] p1r[A29843]DEBSC succinate dehydrogenase (EC 1.3.99.1) cytochrome 558 - Bacillus subtilis	78	57	657
199	4	7209	5557	gi 142521	deoxyribodipyrimidine photolyase [Bacillus subtilis] p1r[AJ7192]AJ7192 uvrB protein - Bacillus subtilis sp 914951 UVRC_BACSU EXCINUCLEASE ABC SUBUNIT C.	78	62	1653
223	3	3831	3523	gi 139596	[Escherichia coli 15200 insertion sequence from ECOR63, partial.], ene product [Escherichia coli]	78	47	309

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
299	4	1865	2149	gi 467439	temperature sensitive cell division [Bacillus subtilis]	78	62	285
321	9	7734	7315	gi 1142879	ORF3 is homologous to an ORF downstream of the spo7 gene of S.coli; RPS [Bacillus stearothermophilus]	78	55	420
352	4	3716	3944	gi 349050	actin 1 (Pneumocystis carinii)	78	42	231
352	5	7592	6093	gi 1903587	NADH dehydrogenase subunit 5 [Bacillus subtilis] ap P39755 NDHFBACSU NADH DEHYDROGENASE SUBUNIT 5 (EC 1.6.5.3) NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5)	78	58	1500
376	1	2	583	gi 551693	dethiobiotin synthase [Bacillus sphaericus]	78	34	582
424	2	1595	1768	gi 1524117	[alpha-acetolactate decarboxylase [Lactococcus lactis]	78	68	174
450	1	1914	988	gi 1030068	NAD(P)H oxidoreductase, isoform reductase homologue [Solanum tuberosum]	78	63	927
558	1	762	562	gi 1513588	bi-functional protein [Methanococcus jannaschii]	78	60	201
670	3	1152	1589	gi 1122759	unknown [Bacillus subtilis]	78	64	438
714	1	64	732	gi 143460	37 kD minor sigma factor (rpoF, sigB; tgg start codon) [Bacillus subtilis]	78	57	669
814	1	3	368	gi 1137783	unknown [Bacillus subtilis]	78	59	366
981	1	1381	692	gi 143802	Gerc2 [Bacillus subtilis]	78	64	690
995	2	978	727	gi 296947	uridine kinase [Escherichia coli]	78	61	252
1045	3	3	401	gi 1407784	orf-1; novel antigen [Staphylococcus aureus]	78	61	399
1163	2	168	186	gi 410117	glutaminolactate decarboxylase [Bacillus subtilis]	78	54	183
2191	1	794	399	gi 215098	excisionase [Bacteriophage 154a]	78	65	396
2933	1	2	181	gi 1204436	pyruvate formate-lyase [Haemophilus influenzae]	78	73	180
3041	2	129	317	gi 624632	GLT [Escherichia coli]	78	53	189
3581	1	105	401	gi 763186	3-ketocacyl-coA thiolase [Saccharomyces cerevisiae]	78	55	297
3709	1	3	230	gi 460689	TVG [Thermotoga maritima]	78	58	228
3974	1	528	265	gi 558839	unknown [Bacillus subtilis]	78	65	264
3980	1	3	401	gi 39956	ITIC [Bacillus subtilis]	78	62	399
4056	1	647	354	gi 1125635	dihydroxy-acid dehydratase [Bacillus subtilis]	78	55	294
4114	1	630	316	gi S09372 S093	hypothetical protein - Trypanosoma brucei	78	62	315
4185	1	3	179	gi 11339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	78	58	177

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4235	1	655	329	gi 558839	unknown [Bacillus subtilis]	78	60	327
4352	1	541	302	gi 603768	HctI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis] gi 603768 HctI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis]	78	63	240
4368	1	612	307	gi 3353678	heavy-metal transporting P-type ATPase (Proceus mirabilis)	78	59	306
4461	1	428	216	gi 1276841	glutamate synthase (GOGAT) [Porphyra purpurea]	78	36	213
4530	1	474	238	gi 39956	ilgic [Bacillus subtilis]	78	65	237
3	2	2969	2073	gi 1109684	ProV [Bacillus subtilis]	77	56	897
12	2	2426	1965	gi 467335	ribosomal protein L9 [Bacillus subtilis]	77	59	462
27	1	2	388	gi 1212728	YqhI [Bacillus subtilis]	77	63	387
39	2	590	1252	gi 40054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) [Bacillus subtilis]	77	60	663
42	6	2704	2931	gi 606241	30S ribosomal subunit protein S14 [Escherichia coli] sp P02370 RS14_BCOU1 30S RIBOSOMAL PROTEIN S14, (S08 2-101)	77	65	228
46	1A	15459	16022	gi 2977798	mitochondrial formate dehydrogenase precursor [Solanum tuberosum] pir JQ2772 JQ2772 formate dehydrogenase (BC 1.2.1.2) precursor, mitochondrial - potato	77	55	1164
100	4	4562	4002	gi 1340128	ORF1 [Staphylococcus aureus]	77	54	561
102	8	5378	5713	gi 1311482	acetolactate synthase [Thermus aquaticus]	77	57	336
109	7	4742	5383	gi 710637	Unknown [Bacillus subtilis]	77	56	642
117	1	2	1228	gi 1237015	ORF4 [Bacillus subtilis]	77	53	1227
124	10	8323	7688	gi 405819	thymidine kinase [Bacillus subtilis]	77	63	836
147	3	1146	985	gi 849027	hypothetical 15.9-kDa protein [Bacillus subtilis]	77	37	162
152	10	7354	7953	gi 1205583	spermidine/putrescine transport ATP-binding protein [Haemophilus influenzae]	77	55	600
169	2	1004	1282	gi 473825	'elongation factor EF-7a' [Escherichia coli]	77	58	279
184	2	380	1147	gi 216314	esterase [Bacillus stearothermophilus]	77	60	768
189	7	3296	3668	gi 853809	ORP3 [Clostridium perfringens]	77	48	573
193	1	132	290	gi 1303788	Yqeh [Bacillus subtilis]	77	54	159
195	8	8740	8414	gi 1499620	M. Jannaschii predicted coding region M20798 [Methanococcus jannaschii]	77	44	327
205	8	5428	5204	gi 216340	ORF for adenylate kinase [Bacillus subtilis]	77	61	225

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
205	29	14795	14502	gi1786155	Ribosomal Protein L23 [Bacillus subtilis]	77	62	294
211	5	1908	2084	gi1410132	OMPX8 [Bacillus subtilis]	77	47	177
217	5	3478	4416	gi1496254	[fibronectin/fibrinogen-binding protein [Streptococcus pyogenes]	77	54	939
222	1	267	998	gi11407784	[orf-1; novel antigen [Staphylococcus aureus]	77	57	732
233	2	1819	1346	gi1467408	unknown [Bacillus subtilis]	77	61	474
243	3	2661	2299	gi1516155	unconventional myosin [Sus scrofa]	77	32	363
299	1	68	769	gi1467436	unknown [Bacillus subtilis]	77	54	702
301	4	1468	1283	gi1950071	ATP-bind. pyrimidine kinase [Mycoplasma capricolum] pir[S48605]S48605 (fragment)	77	48	186
302	5	2741	3211	gi1508980	pheB [Bacillus subtilis]	77	57	471
302	7	3835	4863	gi1147783	ruvB protein [Escherichia coli]	77	60	1029
307	9	5402	4797	gi11070015	[protein-dependent [Bacillus subtilis]	77	60	606
312	1	99	1391	gi1143165	[malic enzyme [EC 1.1.1.38] [Bacillus stearothermophilus] pir[A33307]PERSXS [malate dehydrogenase oxaloacetate-decarboxylating] [EC 1.1.1.38] - Bacillus stearothermophilus	77	62	1293
312	2	1541	2443	gi1139855	[carboxyltransferase beta subunit [Synecococcus RCC7942]	77	58	903
321	5	5666	4596	gi139844	[fumarase (citG) (aa 1-462) [Bacillus subtilis]	77	65	1071
365	1	2	1021	gi1143374	[vmoB [Bacillus subtilis] [phosphoribosyl glycineamide synthetase (PUR-D; gtu start codon) Bacillus subtilis]	77	57	1020
374	1	1	708	gi11405446	[transketolase [Bacillus subtilis]	77	61	708
385	1	1128	565	gi1533099	[endonuclease III [Bacillus subtilis]	77	63	564
392	2	594	1940	gi1556014	[UDP-N-acetyl muramate-alanine ligase [Bacillus subtilis] ap[P40776]KUC_BACSU UDP-N-ACETYLURAMATE--ALANINE LIGASE (EC 3.2.8) [UDP-N- ACETYLURAMATE--L-ALANINE SYNTHETASE] (FRAGMENT)	77	65	1347
405	5	4079	3570	gi11303912	[yqhW [Bacillus subtilis]	77	64	510
487	4	1302	1472	gi1432427	[ORF1 gene product [Acinetobacter calcoaceticus]	77	48	171
522	1	2	562	pir[AD1179]SYNS	[tyrosine--tRNA ligase (EC 6.1.1.1) - Bacillus stearothermophilus	77	63	561

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match Accession	Match (%)	Match gene name	% sim	% ident	Length (nt)
52J	2	1587	1351	gi 1187979		44% identity over 302 residues with hypothetical protein from <i>Synechocystis</i> sp. accession D6406_CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtilis]	77	48	237
536	2	983	612	gi 143366		adenylosuccinate lyase (PUB-B) [Bacillus subtilis] pir C29336 W285DS	77	61	372
548	2	339	872	gi 143387		adenylosuccinate lyase (EC 4.3.2.2) - Bacillus subtilis	77	56	534
559	1	2	481	gi 904198		aspartate transcarbamylase [Bacillus subtilis]	77	33	480
633	2	1747	1313	gi 387577		Hypothetical protein [Bacillus subtilis]	77	64	435
642	1	85	360	gi 46971		ORF1A [Bacillus subtilis]	77	61	276
659	1	125	1219	gi 1072381		lePip gene product [Staphylococcus epidermidis]	77	62	1095
670	4	1587	1620	gi 1122760		glutamyl-aminopeptidase [Lactococcus lactis]	77	58	234
789	1	2	391	gi 1377823		unknown [Bacillus subtilis]	77	65	390
815	1	10	573	gi 1303861		aminopeptidase [Bacillus subtilis]	77	49	584
899	1	1	225	gi 1201444		YggN [Bacillus subtilis]	77	55	225
1083	1	3	188	gi 460828		H. influenzae predicted coding region H10594 [Haemophilus influenzae]	77	66	186
1942	1	415	209	gi 160047		[B959] [Saccharomyces cerevisiae]	77	38	207
2559	1	1	171	gi 1499034		p101/acidic basic repeat antigen (P101) [Plasmodium falciparum] pir A29322 A29322	77	61	171
2933	2	243	401	gi 42370		101K malaria antigen precursor - Plasmodium alciapum (strain Camp)	77	72	159
2965	1	56	292	gi 1524397		M. jamaacii predicted coding region M10255 [Methanococcus jamaacii]	77	45	237
2976	1	614	309	gi 40003		pyruvate formate-lyase (AA 1-760) [Escherichia coli] ix S01788 S01788	77	60	306
2979	2	678	400	gi 1204354		formate C-acetyltransferase (EC 2.3.1.54) - chierichia coli	77	61	279
2988	1	601	377	gi 438465		glycine betaine transporter OpuD [Bacillus subtilis]	77	55	225
2990	1	331	167	gi 142562		oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] p E23129 ODO1_BACSU	77	63	165
3032	1	3	389	gi 468430		2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA-KETOGLUTARATE DEHYDROGENASE)	77	56	387
3057	1	1	195	gi 468764		ATP synthase epsilon subunit [Bacillus megaterium] pir B28599 P0859H	77	50	195
						transporting ATP synthase (EC 3.6.1.34) pailon chain - Bacillus megaterium			
						alcohol dehydrogenase 2 [Entamoeba histolytica]			
						macR gene product [Rhizobium meliloti]			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4008	1	726	400	gi 603768	Hut1 protein, imidazolone-5-propanone hydrolase [Bacillus subtilis]	77	52	327
4048	1	703	386	gi 216278	Gramicidin S synthetase 1 [Bacillus brevis]	77	55	318
4110	1	3	368	pir S52915 S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	366
4115	1	1	348	gi 517205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]	77	65	348
4225	1	590	297	gi 1322245	malonate pyrophosphate decarboxylase [Rattus norvegicus]	77	60	294
4611	2	494	327	gi 508979	GTP-binding protein [Bacillus subtilis]	77	57	166
4668	1	361	182	pir S52915 S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	77	61	180
25	1	2	1627	gi 1150620	HsaA [Streptococcus pneumoniae]	76	58	1626
38	5	1488	2537	pir A43577 A435	regulatory protein p108 - Clostridium perfringens	76	57	1050
52	5	2962	4041	gi 1161061	dioxigenase [Methylobacterium extorquens]	76	62	1080
56	20	27389	27955	gi 467402	unknown [Bacillus subtilis]	76	56	567
57	15	12046	12219	gi 1206040	weak similarity to koratin [Caenorhabditis elegans]	76	40	174
91	2	1062	2261	gi 45715	acetyl coenzyme A acetyltransferase (thiolase) [Clostridium acetobutylicum]	76	57	1200
98	2	818	1624	gi 467422	unknown [Bacillus subtilis]	76	62	807
98	5	2965	3228	gi 897793	ly98 gene product [Pediococcus acidilactici]	76	52	264
98	8	5922	6326	gi 467427	methionyl-tRNA synthetase [Bacillus subtilis]	76	53	405
104	3	1322	1885	gi 216151	RNA polymerase (gene 1; t1g start codon) [Bacteriophage SP02] gi 579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] pir A1498 DJBSP2 DNA-directed DNA polymerase (EC 2.7.7) - phage P02	76	63	564
124	9	8134	7055	gi 853776	peptide chain release factor 1 [Bacillus subtilis] pir S55437 S55437 peptide chain release factor 1 - Bacillus subtilis	76	58	1080
164	5	2832	3311	gi 1204976	prolyl-tRNA synthetase [Haemophilus influenzae]	76	53	480
168	2	2617	1841	gi 1177253	putative ATP-binding protein of ABC-type [Bacillus subtilis]	76	58	177
189	2	163	888	gi 467384	unknown [Bacillus subtilis]	76	63	726
225	3	2253	3518	gi 142936	[poly]-polyglutamate synthetase [Bacillus subtilis] pir B04646 B04646 folC - Bacillus subtilis	76	53	1266
236	1	335	925	gi 1146197	putative [Bacillus subtilis]	76	54	591
237	8	5323	5541	gi 1279261	PI3G.6 [Caenorhabditis elegans]	76	47	219

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
263	5	5490	4585	gi 1510348	[dihydrodipicolinate synthase (Methanococcus jannaschii)]	76	49	906
304	3	1051	1794	gi 666982	[putative membrane spanning subunit (Bacillus subtilis) pif S52382 S52382 probable membrane spanning protein - Bacillus subtilis]	76	60	744
332	4	3611	6624	gi 143312	[6-phospho-1-fructokinase (gtg start codon; EC 2.7.1.11) (Bacillus thermoautotrophicus)]	76	56	1014
343	1	2	1036	gi 405956	[yaeE (Escherichia coli)]	76	59	1035
347	1	409	1701	gi 396304	[acetylornithine decarboxylase (Escherichia coli)]	76	72	1293
358	1	672	1907	gi 1146215	[39.0% identity to the Escherichia coli S1 ribosomal protein; putative (Bacillus subtilis)]	76	58	1236
371	1	1	222	gi 537084	[alternate gene name mgt; CG Site No. 497 (Escherichia coli) pif S5468 S5468 mgtA protein - Escherichia coli]	76	61	222
379	4	4331	4858	gi 143268	[dihydrodipicolinate transsuccinylase (odhB; EC 2.3.1.61) (Bacillus subtilis)]	76	61	528
404	5	4022	4492	gi 1303823	[yqfG (Bacillus subtilis)]	76	60	471
411	1	2	307	gi 186025	[ORF YKL027w (Saccharomyces cerevisiae)]	76	55	306
472	3	4356	2854	gi 1405464	[AlST (Bacillus subtilis)]	76	57	1503
546	1	273	995	gi 153821	[streptococcal pyrogenic exotoxin type C (speC) precursor Streptococcus pyogenes]	76	36	723
588	1	1054	557	gi 1002520	[MutS (Bacillus subtilis)]	76	61	498
591	1	16	735	gi 885934	[ClpB (Synecococcus sp.)]	76	44	720
602	2	175	798	gi 1486422	[OppD homologue (Rhizobium sp.)]	76	52	624
619	2	547	290	gi 330613	[major capsid protein (Human cytomegalovirus)]	76	47	258
660	4	2568	3302	gi 904199	[hypothetical protein (Bacillus subtilis)]	76	55	735
677	1	452	228	gi 40177	[spoOF gene product (Bacillus subtilis)]	76	58	225
962	1	24	206	gi 142443	[adenylosuccinate synthetase (Bacillus subtilis) sp P29726 PUBA_BACSU ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) IMP-ASPARTATE LIGASE)]	76	67	183
978	1	1158	580	gi 1511333	[M. jannaschii predicted coding region M1322 (Methanococcus jannaschii)]	76	56	579
997	1	486	244	gi 467154	[No definition line found (Mycobacterium leprae)]	76	38	243
1563	1	529	266	gi 1303984	[yqkO (Bacillus subtilis)]	76	52	264
2184	1	361	182	gi 506706	[CapJ (Staphylococcus aureus)]	76	38	180
3572	1	1	387	gi 153898	[transport protein (Salmonella typhimurium)]	76	65	387

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	seqch gene name	% sim	% ident	length (nt)
2942	1	29	400	gi 710020	nitrite reductase (nirB) [Bacillus subtilis]	76	59	372
2957	1	377	216	gi 1511251	hypothetical protein (SP:942404) [Methanococcus jannaschii]	76	47	162
2980	1	554	279	gi 1403464	AlaT [Bacillus subtilis]	76	53	276
3015	1	649	326	gi 408115	ornithine acetyltransferase [Bacillus subtilis]	76	61	324
3124	1	13	174	gi 882705	ORF_0401 [Escherichia coli]	76	65	162
3179	1	3	161	gi 1168477	ferredoxin-dependent glutamate synthase [Zea mays] p1r[A38596]A38596 glutamate synthase (ferredoxin) (SC 1.4.7.1) - size	76	53	159
3789	1	2	379	gi 39956	ITIGC [Bacillus subtilis]	76	55	378
3892	1	3	314	gi 1510398	ferripyochelin binding protein [Methanococcus jannaschii]	76	52	312
3928	1	798	400	gi 143016	permease [Bacillus subtilis]	76	59	399
4159	1	757	386	sp P80544 MRSP_	METHICILLIN-RESISTANT SURFACE PROTEIN (FRAGMENTS)	76	66	372
4204	1	17	331	gi 296464	ATPase [Lactococcus lactis]	76	56	315
4394	1	494	249	gi 987255	Menkes disease gene (Homo sapiens)	76	48	246
4506	1	2	313	gi 216746	D-lactate dehydrogenase (Lactobacillus plantarum)	76	47	312
4546	1	477	247	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	76	61	231
4596	1	379	191	gi 560027	cellulose synthase [Acetobacter xylinum]	76	70	189
4	5	5257	4337	gi 882532	ORF_0294 [Escherichia coli]	75	59	921
6	1	164	952	gi 40960	OTCase [Escherichia coli]	75	56	789
12	3	5935	3944	gi 467336	unknown [Bacillus subtilis]	75	57	1992
21	18	18272	17310	gi 1296433	O-acetylserine sulphydrylase B [Alcaligenes eutrophus]	75	55	963
25	3	2356	3393	gi 1502419	Plax [Bacillus subtilis]	75	56	1038
36	8	5765	6037	gi 1256517	unknown [Schizosaccharomyces pombe]	75	45	273
46	13	11186	12058	gi 48972	nitrate transporter [Synechococcus sp.]	75	46	873
51	7	2474	3677	gi 143607	sporulation protein [Bacillus subtilis]	75	61	204
53	16	14850	16590	gi 143402	recombination protein (csg start codon) [Bacillus subtilis] gi J03923 RecW [Bacillus subtilis]	75	51	261
74	3	3572	2568	gi 1204847	ornithine carbamoyltransferase [Haemophilus influenzae]	75	61	1005

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match	match gene name	% sim	% ident	length (nt)
85	3	4628	3930	gi 143368		phosphoribosylformyl glycylamide synthetase I (PUR-L; gta start odon) [Bacillus subtilis]	75	63	699
85	5	5588	4878	gi 143367		phosphoribosyl aminodazole succinocarboxamide synthetase (PUR-C; gta start codon) [Bacillus subtilis]	75	55	711
85	8	6625	7330	gi 1303916		Yqia [Bacillus subtilis]	75	53	906
87	3	2340	3590	gi 1064813		homologous to sp:phoA_BACSU [Bacillus subtilis]	75	56	1251
87	6	6084	6896	gi 1064810		function unknown [Bacillus subtilis]	75	61	813
108	2	1844	1503	gi 1001824		hypothetical protein [Synecococcus sp.]	75	51	342
110	3	1748	3727	gi 1147593		putative ptop synthetase [Streptomyces coelicolor]	75	55	1980
110	7	4151	5252	gi 1177251		clwD gene product [Bacillus subtilis]	75	75	900
120	14	11266	10649	gi 1524394		ORF-2 upstream of gbaB operon [Bacillus subtilis]	75	55	618
121	5	2050	4221	gi 1154632		Nidg [Bacillus subtilis]	75	54	2172
124	1	283	143	gi 405622		unknown [Bacillus subtilis]	75	56	141
128	1	81	1139	gi 143316		[gen] gene products [Bacillus megaterium]	75	48	1059
130	8	5760	5903	gi 1256654		54.8% identity with Neisseria gonorrhoeae regulatory protein PilB; putative [Bacillus subtilis]	75	62	144
136	2	4480	3185	gi 467403		acyl-CoA synthetase [Bacillus subtilis]	75	54	1296
161	10	5439	5798	gi 1001195		hypothetical protein [Synecococcus sp.]	75	55	360
172	4	3819	2995	gi 755153		ATP-binding protein [Bacillus subtilis]	75	52	825
179	1	2024	1107	gi 143037		porphobilinogen deaminase [Bacillus subtilis]	75	58	918
195	10	9529	9174	sp P25745 YCPD_		HYPOTHETICAL PROTEIN IN PURB 5'-REGION (ORF-15) (FRAGMENT)	75	60	156
200	4	2605	4596	gi 142440		ATP-dependent nuclease [Bacillus subtilis]	75	56	1992
206	3	6900	5620	gi 1256135		YdbP [Bacillus subtilis]	75	53	1281
216	2	159	389	gi 1052800		unknown [Schizosaccharomyces pombe]	75	58	231
229	1	29	847	gi 1205958		branched chain aa transport system II carrier protein (Haemophilus influenzae)	75	49	819
230	2	518	1714	gi 971337		nitrite extrusion protein [Bacillus subtilis]	75	53	1197
231	1	2240	1122	gi 1002521		MutL [Bacillus subtilis]	75	54	1119
233	3	1314	1859	gi 467405		unknown [Bacillus subtilis]	75	59	546

TABLE 2

3. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
289	1	325	164	gi11511246	[methyl coenzyme M reductase system, component A2 [Methanococcus jannaschii]]	75	50	162
292	1	1389	772	gi11511604	[M. jannaschii predicted coding region MJ1651 [Methanococcus jannaschii]]	75	46	618
304	4	1773	2261	gi1205328	[surfactin [Haemophilus influenzae]]	75	55	489
312	3	2437	3387	gi1285621	[undefined open reading frame [Bacillus stearothermophilus]]	75	62	951
312	5	4622	6403	gi11041097	[Pyruvate Kinase [Bacillus psychrophilus]]	75	57	1782
319	1	353	877	gi11212728	[yqhi [Bacillus subtilis]]	75	54	525
320	5	4321	5031	gi11070361	[OMP decarboxylase [Lactococcus lactis]]	75	56	711
320	6	5010	5642	gi1143394	[OMP-PRPP transferase [Bacillus subtilis]]	75	60	633
337	4	1519	2088	gi1487433	[citrate synthase II [Bacillus subtilis]]	75	58	570
394	2	669	1271	gi1304976	[matches P500017: ATP-GTP_A and P500301: EFACITOR_GTP; similar to longation factor G, TetM/TetO tetracycline-resistance proteins Escherichia coli]	75	51	603
423	1	127	570	gi11183839	[unknown [Pseudomonas aeruginosa]]	75	59	444
433	2	1603	1929	gi1149211	[acetolactate synthase [Klebsiella pneumoniae]]	75	63	327
446	2	176	1540	gi1312441	[dihydroorotase [Bacillus caldolyticus]]	75	62	1365
486	1	494	249	gi11149682	[potF gene product [Clostridium perfringens]]	75	55	246
496	1	3	794	gi1143582	[spoIIIEA protein [Bacillus subtilis]]	75	59	792
498	2	824	1504	gi1143328	[phoP protein (put.) putative [Bacillus subtilis]]	75	47	681
499	2	1061	1624	gi11387979	[44% identity over 302 residues with hypothetical protein from Synchocystis sp. accession D54006.CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices [Bacillus subtilis]]	75	51	564
568	1	641	453	gi11304110	[triacylglycerol lipase (EC 3.1.1.3) 2 - Mycoplasma mycoides subsp. mycoides (SC3)]	75	50	189
613	2	430	233	gi1330993	[tegument protein [Saimaeline herpesvirus 2]]	75	75	198
623	1	1	525	gi1529754	[apeC [Streptococcus pyogenes]]	75	43	525
642	5	1809	2474	gi11176401	[EpiG [Staphylococcus epidermidis]]	75	51	666
646	2	454	657	gi1172442	[ribonuclease P [Saccharomyces cerevisiae]]	75	37	204
657	1	3	347	gi1882541	[omf_0236 [Escherichia coli]]	75	47	345
750	1	1662	832	gi146971	[epiP gene product [Staphylococcus epidermidis]]	75	57	831

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match	match	match gene name	% sim	% ident	length (nt)
754	1	2	481	gi 103901	YcbT [Bacillus subtilis]		75	57	480
763	2	563	393	gi 1205145	multidrug resistance protein [Haemophilus influenzae]		75	51	171
775	1	561	482	pir 036889 D368	leuA protein, inactive - Lactococcus lactis subsp. lactis (strain IL1403)		75	63	480
793	1	1	180	gi 143316	[gap] gene products [Bacillus megaterium]		75	57	180
800	1	318	160	gi 509411	NFRA protein [Azorhizobium caulinodans]		75	34	159
811	1	1117	560	gi 143434	Rho factor [Bacillus subtilis]		75	60	558
940	1	493	329	gi 1276985	arginase [Bacillus caldoveloxi]		75	50	165
971	2	37	352	gi 1001373	hypothetical protein [Synecocystis sp.]		75	58	216
1059	1	384	232	gi 726480	L-glutamine-D-fructose-6-phosphate amidotransferase [Bacillus subtilis]		75	67	153
1109	2	219	374	gi 143331	alkaline phosphatase regulatory protein phoR - Bacillus subtilis pir A27650 A27650 regulatory protein phoR - Bacillus subtilis pir P23545 PHO_R_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHO_R (EC 2.7.3.-)		75	53	156
1268	1	271	137	gi 104135	ornithine acetyltransferase [Bacillus stearothermophilus] spi Q07908 ARGJ_BACST GLUTAMATE N-ACETYLTRANSFERASE (EC 2.3.1.35) ORNITHINE ACETYLTRANSFERASE [ORNITHINE TRANSACETYLASE] (OATASE) / HIND-ACID ACETYLTRANSFERASE (EC 2.3.1.1) (N-ACETYLGLUTAMATE YNTIIA		75	63	135
1500	1	324	163	gi 1205488	excinuclease ABC subunit B [Haemophilus influenzae]		75	57	162
1529	1	798	400	gi 1002521	MutL [Bacillus subtilis]		75	54	399
3010	1	770	387	gi 1204435	pyruvate formate-lyase activating enzyme [Haemophilus influenzae]		75	54	384
3105	1	1	180	gi 1041097	Pyruvate Kinase [Bacillus psychrophilus]		75	57	180
3117	1	45	212	gi 859317	peptide synthetase module [Microcystis aeruginosa] pir S49111 S49111 probable amino acid activating domain - Microcystis aeruginosa (fragment) (SUA 144-528)		75	42	168
3119	2	139	345	gi 145294	adenine phosphoribosyl-transferase [Bacterichia coli]		75	66	207
3880	1	618	310	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]		75	58	309
3911	1	48	401	gi 433991	ATP synthase subunit beta [Bacillus subtilis]		75	68	354
3957	1	2	379	pir D16889 D368	3-isopropylmalate dehydratase (EC 4.2.1.33) chain leuC - Lactococcus lactis subsp. lactis (strain IL1403)		75	65	378
4005	1	5	259	gi 216746	D-lactate dehydrogenase [Lactobacillus plantarum]		75	48	255
4080	1	73	333	gi 415855	deoxyribose aldolase [Mycoplasma hominis]		75	59	261

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Cunfig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4111	1	1	339	gi 169435	putative [Lactococcus lactis]	75	57	339
4136	1	602	303	gi 450688	hslM gene of Ecopri gene product [Escherichia coli] pif[S3637]S3637 hsdM protein - Escherichia coli pif[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	75	56	300
4144	1	669	336	gi 148972	nitrate transporter [Synecoccus sp.]	75	49	333
4237	1	664	374	gi 1339950	large subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	75	55	291
4306	2	73	318	gi 294260	major surface glycoprotein [Pneumocystis carinii]	75	68	246
4343	1	715	359	gi 1204652	methylated-DNA--protein-cysteine methyltransferase [Haemophilus influenzae]	75	52	357
4552	1	620	312	gi 296464	ATPase [Lactococcus lactis]	75	55	309
48	9	5776	6126	gi 443793	HupC [Escherichia coli]	74	50	351
50	8	6910	6221	gi 1239988	hypothetical protein [Bacillus subtilis]	74	55	690
56	9	10770	12221	gi 1000451	TreP [Bacillus subtilis]	74	57	1452
64	2	2266	1622	gi 41015	aspartate-2-oxoglutarate lyase [Escherichia coli]	74	57	645
66	6	5063	4848	gi 1212729	YqhJ [Bacillus subtilis]	74	47	216
67	18	14334	16897	gi 1310631	endoglucanase [Methanococcus jannaschii]	74	52	564
102	15	12561	13136	gi 149429	putative [Lactococcus lactis]	74	67	576
102	16	13121	14419	gi 149435	putative [Lactococcus lactis]	74	57	1299
108	4	4873	3902	gi 39478	ATP binding protein of transport ATPases [Bacillus firmus] ir[S15406]S15406 ATP-binding protein - Bacillus firmus p[P28946]P28946 HYPOTHETICAL ATP-BINDING TRANSPORT PROTEIN	74	59	972
116	5	8574	7093	gi 1205430	dipeptide transport system permease protein [Haemophilus influenzae]	74	49	1482
120	7	4342	4803	gi 146970	ribonucleoside triphosphate reductase [Escherichia coli] pif[A47331]A47331 anaerobic ribonucleotide reductase - Escherichia coli	74	58	462
121	7	5961	6581	gi 1107528	lctg start [Campylobacter coli]	74	51	621
128	3	2320	3531	gi 143318	phosphoglycerate kinase [Bacillus megaterium]	74	57	1212
130	7	5237	5791	gi 1256653	DNA-binding protein [Bacillus subtilis]	74	60	555
136	3	6745	5150	gi 143076	histidase [Bacillus subtilis]	74	58	1596
145	2	664	1368	gi 407773	devA gene product [Anabaena sp.]	74	45	705
152	1	552	277	gi 1377833	unknown [Bacillus subtilis]	74	54	276

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	saajq gene name	% sim	% ident	length (nt)
164	10	11064	11375	gi1580900	ORF3 gene product [Bacillus subtilis]	74	52	312
175	2	3109	2624	gi1642636	unknown [Rhizobium meliloti]	74	34	486
175	9	6064	5612	gi1854656	Na/H antiporter system ORF2 [Bacillus alcalophilus]	74	46	453
195	11	11346	10339	gi1204430	hypothetical protein (SP:25745) [Haemophilus influenzae]	74	55	1008
205	17	9619	9059	gi11044979	ribosomal protein L6 [Bacillus subtilis]	74	64	561
236	7	5574	6710	gi1146207	putative [Bacillus subtilis]	74	63	1137
241	3	4521	3334	gi1694121	malate thiokinase [Methylobacterium extorquens]	74	52	1188
246	6	3105	2799	gi1467374	single strand DNA binding protein [Bacillus subtilis]	74	64	507
249	4	6551	5313	gi11524397	glycine betaine transporter Opd [Bacillus subtilis]	74	55	1239
261	7	4389	4081	gi1809542	CbrB protein [Brwinia chrysanthemi]	74	42	309
278	6	5714	4665	gi1204072	ATP-binding protein [Haemophilus influenzae]	74	54	1050
309	1	1220	666	gi1205579	hypothetical protein (GB:U14003_302) [Haemophilus influenzae]	74	53	555
315	2	1473	862	gi1143398	guinol oxidase [Bacillus subtilis]	74	57	612
320	1	1	1065	gi1943389	glutaminase of carbamyl phosphate synthetase [Bacillus subtilis] pir[E39845]E39845 carbamoyl-phosphate synthase glutamine-hydrolysing [EC 6.3.5.5], pyridine-repressible, small heat - Bacillus subtilis	74	60	1065
340	2	382	1126	gi1534857	ATPase subunit a [Bacillus stearothermophilus]	74	56	747
405	2	1742	1311	gi1303915	VqH2 [Bacillus subtilis]	74	65	432
433	5	2503	3270	gi1473902	alpha-acetolactate synthase [Lactococcus lactis]	74	56	768
452	1	1	942	gi1413982	lpa-58F gene product [Bacillus subtilis]	74	52	942
461	1	3	1193	gi1558494	homoserine dehydrogenase [Bacillus subtilis]	74	51	1191
461	2	1174	1407	gi140211	threonine synthase (thtC) (AA 1-352) [Bacillus subtilis] ir[A25364]A25364 threonine synthase [EC 4.2.99.2] - Bacillus btillis	74	56	234
462	2	402	734	gi1162520	thioredoxin [Bacillus subtilis]	74	62	333
478	1	574	320	gi11699005	glycyl-tRNA synthetase [Methanococcus jannaschii]	74	52	255
501	2	739	1740	gi1217040	acid glycoprotein [Streptococcus pyogenes]	74	58	1002
551	2	4083	2791	gi143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir[D42728]D42728 glutamate-1-semialdehyde 2,1-aminomutase [EC 4.3.8] - Bacillus subtilis	74	51	1293

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
573	1	1	477	gi 1006605	hypothetical protein [Synchocystis sp.]	74	45	477
596	2	1780	1298	gi 1303853	YqgP [Bacillus subtilis]	74	55	483
618	2	2924	1758	gi 1146237	21.4% of identity to trans-acting transcription factor of <i>Saccharomyces cerevisiae</i> ; 25% of identity to sucrose synthase of <i>Zea mays</i> ; putative [Bacillus subtilis]	74	55	1167
659	2	1269	1595	gi 1072380	ORF3 [Lactococcus lactis]	74	62	327
724	1	373	188	gi 143374	phosphoribosyl glycineamide synthetase (PUR-D, gtp start codon) [Bacillus subtilis]	74	58	186
743	2	604	1209	gi 153833	ORF1; putative [Streptococcus parasanguis]	74	50	606
836	1	2	259	gi 143458	ORF V [Bacillus subtilis]	74	47	258
989	2	443	724	gi 1303994	YqkM [Bacillus subtilis]	74	46	282
1106	1	1	492	gi 146970	lepD gene product [Staphylococcus epidermidis]	74	54	492
1135	2	373	528	gi 413948	ipa-24d gene product [Bacillus subtilis]	74	48	156
1234	1	817	452	gi 495245	rocJ gene product [Erwinia chrysanthemi]	74	36	366
2586	1	2	238	gi 1149701	sbcc gene product [Clostridium perfringens]	74	62	237
2959	1	798	400	gi 1405454	aconitase [Bacillus subtilis]	74	60	399
2962	1	650	363	gi 450686	3-phosphoglycerate kinase [Thermotoga maritima]	74	58	288
2983	1	3	191	gi 1303893	YqkL [Bacillus subtilis]	74	56	189
3018	1	2	223	gi 143040	glutamate-1-semialdehyde 2,1-aminotransferase [Bacillus subtilis] pir D42728 D42728 glutamate-1-semialdehyde 2,1-aminotransferase (EC 4.3.8) - [Bacillus subtilis]	74	56	222
3038	1	510	256	pir 552915 5529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	74	57	255
3062	1	374	189	gi 1107528	ltaG start [Campylobacter coli]	74	51	186
4035	1	184	360	gi 1022725	unknown [Staphylococcus haemolyticus]	74	66	177
4045	1	607	305	gi 1510977	M. jannaschii predicted coding region M30938 [Methanococcus jannaschii]	74	41	303
4283	1	471	304	gi 520844	orf4 [Bacillus subtilis]	74	58	168
4445	1	3	221	gi 580910	peptide-synthetase ORF1 [Bacillus subtilis]	74	54	219
4587	1	458	231	gi 1370207	orf6 [Lactobacillus sakei]	74	59	238

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
4603	1	29	214	gi 146208	glutamate synthase large subunit (EC 2.6.1.53) [Escherichia coli] pfr A29617 A29617 glutamate synthase (NADPH) (EC 1.4.1.13) large hain - Escherichia coli	74	60	186
4670	1	366	184	gi 256135	YbbP [Bacillus subtilis]	74	61	183
5	10	7953	7162	gi 143727	putative [Bacillus subtilis]	73	42	792
11	2	2454	1372	gi 166338	dihydroacetate dehydrogenase [Agrobacterium tumefaciens]	73	55	1083
14	1	2024	1020	gi 143373	phosphoribosyl aminimidazole carboxy formyl ornyltransferase/inosine monophosphate cyclohydrolyase [Pur-H(2)] [Bacillus subtilis]	73	54	1005
23	5	5426	4635	gi 1468939	meso-2,3-butenediol dehydrogenase (D-acetoin forming) (Klebsiella pneumoniae)	73	58	792
23	17	17379	16360	gi 297060	ornithine cyclodeaminase [Rhizobium meliloti]	73	37	1020
29	2	692	1273	gi 467442	stage V sporulation [Bacillus subtilis]	73	54	582
31	5	6467	4914	gi 414000	lipa-76d gene product [Bacillus subtilis]	73	55	1554
37	8	8658	7402	gi 429259	pepr gene product [Bacillus subtilis]	73	59	1257
37	9	7738	7562	gi 168367	alpha-isopropylmalate isomerase (put.); putative [Rhizomucor ircinellolides]	73	52	177
38	7	3931	4896	gi 405885	lysN [Escherichia coli]	73	58	966
44	6	5041	4238	gi 580895	unknown [Bacillus subtilis]	73	53	804
44	11	7767	8306	gi 42009	maab gene product [Escherichia coli]	73	50	540
45	3	2439	3080	gi 1109685	ProW [Bacillus subtilis]	73	47	642
54	13	14036	13794	gi 413931	lipa-7d gene product [Bacillus subtilis]	73	61	243
59	4	1430	2248	gi 147923	threonine dehydratase 2 (EC 4.2.1.16) [Escherichia coli]	73	53	819
65	1	1458	730	gi 677944	AppF [Bacillus subtilis]	73	56	729
80	2	1375	860	gi 580932	murD gene product [Bacillus subtilis]	73	53	516
102	13	10124	11179	gi 580891	3-isopropylmalate dehydrogenase (AA 1 - 365) [Bacillus subtilis] pfr A26522 A26522 3-isopropylmalate dehydrogenase (EC 1.1.1.85) - acillus subtilis	73	55	1056
109	2	3493	2600	gi 1510849	M. jannaschii predicted coding region M0775 [Methanococcus jannaschii]	73	40	894
120	8	4782	5756	gi 146970	ribonucleoside triphosphate reductase [Escherichia coli] pfr A47331 A47331 anaerobic ribonucleotide reductase - Escherichia coli	73	56	975
120	9	5726	6223	gi 1204333	anaerobic ribonucleoside-triphosphate reductase [Haemophilus influenzae]	73	62	498

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
132	5	4151	4363	gi 871046	HSP92 - heavy chain potential motor protein [Giardia intestinalis]	73	43	213
140	6	5952	4324	gi 834107	kdpB [Escherichia coli]	73	59	1629
142	6	7060	5919	gi 410125	rflU gene product [Bacillus subtilis]	73	57	1122
149	4	1866	1717	gi 460892	heparin binding protein-44, HBP-44 (mice, Peptide, 360 aa) p1r-JX0281/JX0281 heparin-binding protein-44 precursor - mouse gi 220434 ORF [Mus musculus] (SUB 2-360)	73	53	150
156	1	1	1431	gi 882504	ORF f560 [Escherichia coli]	73	57	1431
174	6	5352	4525	gi 1146240	kutopentolate hydromethyltransferase [Bacillus subtilis]	73	55	828
175	8	5537	5178	gi 854657	Na/H antiporter system ORF3 [Bacillus alcalophilus]	73	56	360
186	5	6593	5493	gi 467477	unknown [Bacillus subtilis]	73	48	1101
249	6	6283	5729	gi 1524397	glycine betaine transporter Opud [Bacillus subtilis]	73	56	555
265	4	1873	2280	gi 139848	U3 [Bacillus subtilis]	73	41	408
270	1	328	582	gi 780461	220 kDa polypeptide (African swine fever virus)	73	53	255
278	4	4283	3618	gi 1208965	hypothetical 23.3 kD protein [Escherichia coli]	73	49	666
279	3	4984	3593	gi 1185288	isochorismate synthase [Bacillus subtilis]	73	58	1392
291	4	1207	1575	gi 1511440	glutamine--fructose-6-phosphate transaminase [Methanococcus jannaschii]	73	63	369
299	2	735	1166	gi 467437	unknown [Bacillus subtilis]	73	58	432
299	5	2050	3234	gi 467439	temperature sensitive cell division [Bacillus subtilis]	73	53	1185
334	1	1237	728	gi 536655	ORF YB244W [Saccharomyces cerevisiae]	73	43	510
336	2	1827	1036	gi 790943	urea amidolyase [Bacillus subtilis]	73	51	792
374	3	1389	1874	gi 405451	YneJ [Bacillus subtilis]	73	55	486
433	4	1916	2554	gi 473902	alpha-acetolactate synthase (Lactococcus lactis)	73	34	639
509	2	1795	1028	gi 467483	unknown [Bacillus subtilis]	73	56	768
513	1	1709	918	gi 1146220	NAD+ dependent glycerol-3-phosphate dehydrogenase [Bacillus subtilis]	73	56	792
533	2	239	733	gi 1510605	hypothetical protein (SP:42297) [Methanococcus jannaschii]	73	44	495
546	2	1148	2815	gi 41748	hsmM protein (AA 1-520) [Escherichia coli]	73	52	1668
549	1	762	382	gi 1314847	CinA [Bacillus subtilis]	73	57	381
567	1	1346	675	gi 410137	ORFX13 [Bacillus subtilis]	73	58	672

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
716	2	654	1112	gi 1256623	exodeoxyribonuclease [Bacillus subtilis]	73	56	459
772	1	3	677	gi 142010	Shows 70.2% similarity and 48.6% identity to the EnvM protein of <i>Almonella typhimurium</i> [Anebaena sp.]	73	57	675
774	1	3	209	gi 409286	bmrU [Bacillus subtilis]	73	52	207
782	1	1	402	gi 143320	[gap] gene products [Bacillus megaterium]	73	56	402
789	2	451	762	gi 1063246	low homology to P14 protein of Hemophilus influenzae and 14.2 kDa protein of <i>Escherichia coli</i> [Bacillus subtilis]	73	56	312
796	1	3	911	gi 453754	ABC transporter [Bacillus subtilis]	73	58	909
806	3	1209	949	gi 143786	cryptophanyl-tRNA synthetase [EC 6.1.1.2] [Bacillus subtilis] pir J70481 YMB5 cryptophan-tRNA ligase [EC 6.1.1.2] - <i>Bacillus subtilis</i>	73	51	261
816	2	4839	3097	gi 41748	hcdH protein (AA 1-520) [Escherichia coli]	73	52	1743
839	1	798	400	gi 886906	argininosuccinate synthetase [Streptomyces clavuligerus] pir S57659 S57659 argininosuccinate synthase [EC 6.3.4.5] - <i>Streptomyces clavuligerus</i>	73	59	399
857	1	3	290	gi 348052	acetoin utilization protein [Bacillus subtilis]	73	50	288
1008	1	790	398	gi 40100	rodC (tag3) polypeptide (AA 1-746) [Bacillus subtilis] lr S06049 S06049 rodC protein - <i>Bacillus subtilis</i> plp3485 TAGP_BACSU TECHOIC ACID BIOSYNTHESIS PROTEIN F.	73	41	393
1018	1	1	213	gi 529357	No definition line found (Caenorhabditis elegans) sp P46975 STT3_CABEL OLIGOSACCHARYL TRANSFERASE STT3 SUBUNIT OHOLOG.	73	53	213
1013	1	3	491	gi 142706	comG1 gene product [Bacillus subtilis]	73	51	489
1174	1	395	204	gi 1149513	alpha2 subunit of laminin 5 [Homo sapiens]	73	60	192
1175	1	655	329	gi 473817	'ORF' [Escherichia coli]	73	57	327
1187	1	3	209	gi 580870	ipa-37d goxA gene product [Bacillus subtilis]	73	52	207
1206	1	72	245	gi 144816	formyltetrahydrofolate synthetase (FHFS) (ttg start codon) [EC 3.4.3] [Moorella thermoacetica]	73	43	174
1454	1	423	241	gi 1213253	unknown [Schizosaccharomyces pombe]	73	53	183
1469	1	517	260	gi 1303787	YqeG [Bacillus subtilis]	73	55	258
1761	1	374	189	gi 9135	Mat26A gene product [Drosophila simulans]	73	34	186
1849	1	467	243	gi 162307	DNA topoisomerase II [Trypanosoma cruzi]	73	60	225
2055	1	2	400	gi 59381	P47K protein [Rhodococcus erythropolis]	73	34	399
2556	1	2	244	gi 145925	fecB [Escherichia coli]	73	62	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
2947	2	549	400	gi 1184680	[polynucleotide phosphorylase (Bacillus subtilis)]	73	51	150
2956	1	746	375	gi 42397	[quinol oxidase (Bacillus subtilis)]	73	58	372
3037	1	655	329	gi 111091	[acetolactate synthase (Bacillus subtilis)]	73	55	327
3115	1	385	194	gi 323866	[overlapping out-of-phase protein (Egplant mosaic virus) sp P20129 V70K_EPMV 70 MD PROTEIN.	73	53	192
3603	2	700	527	gi 1419521	[glutaryl-CoA dehydrogenase precursor (Mus musculus)]	73	48	174
3743	1	798	400	gi 450689	[hcdH gene of EcoPCR1 gene product (Escherichia coli) pIR[S38437/S38437 hcdH protein - Escherichia coli pIR[S09629/S09629 hypothetical protein A - Escherichia coli (SUB 40-520)]	73	54	399
3752	1	640	359	gi 1524193	[unknown (Mycobacterium tuberculosis)]	73	59	282
3852	1	2	181	gi 216746	[D-lactate dehydrogenase (Lactobacillus plantarum)]	73	68	180
3914	1	475	239	gi S13490/S134	[Hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Chicken (fragment)]	73	53	237
3914	2	570	343	gi 528991	[unknown (Bacillus subtilis)]	73	38	228
4069	1	2	316	gi 40003	[oxoglutarate dehydrogenase (NADP+)] (Bacillus subtilis) pP33129 D001_BACSU 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT [EC 2.4.2] (ALPIIA - KETOGLUTARATE DEHYDROGENASE)	73	55	315
4165	1	715	365	gi 1419521	[glutaryl-CoA dehydrogenase precursor (Mus musculus)]	73	48	351
4196	1	1	177	gi 809660	[deoxyribose-phosphate aldolase (Bacillus subtilis) pIR[S49455/S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - actillus subtilis]	73	60	177
4202	1	572	378	gi 528991	[unknown (Bacillus subtilis)]	73	38	195
4314	1	2	193	gi 436797	[N-acyl-L-amino acid amidohydrolase (Bacillus stearothermophilus) sp P3112 AMA_BACST N-ACYL-L-AMINO ACID AMIDOHYDROLASE (EC 5.1.14) (AMINOACYLASE)]	73	47	192
4393	1	3	263	gi 216267	[ORF2 (Bacillus megaterium)]	73	47	261
35	2	903	1973	gi 1146196	[phosphoglycerate dehydrogenase (Bacillus subtilis)]	72	53	1071
38	22	19094	17877	gi 602031	[similar to trimethylamine DH (Mycoplama capricolum) pIR[S49950/S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SGC3) (fragment)]	72	54	1218
38	23	18134	19162	gi 413968	[lpa-44d gene product (Bacillus subtilis)]	72	54	1029
44	19	11895	12953	gi 516272	[unknown (Bacillus subtilis)]	72	49	1059
48	7	6248	7117	gi 434699	[pyruvate synthase (Halobacterium halobium)]	72	49	870
50	7	6563	5691	gi 1205399	[proton glutamate symport protein (Haemophilus influenzae)]	72	53	873

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
53	9	10521	9259	gi 1303956	Y01G [Bacillus subtilis]	72	52	1263
56	23	29569	29995	gi 467471	unknown [Bacillus subtilis]	72	47	447
69	4	5298	4123	gi 1354775	pfoS/R [Treponema pallidum]	72	46	1176
69	5	4377	4982	gi 904198	hypothetical protein [Bacillus subtilis]	72	43	606
73	1	2	856	gi 142997	glycerol uptake facilitator [Bacillus subtilis]	72	59	855
98	13	9371	10258	gi 467435	unknown [Bacillus subtilis]	72	50	888
127	1	1	1593	gi 217144	alanine carrier protein [thermophilic bacterium PS3] pir[A45111][A45111 alanine transport protein - thermophilic acterium PS-3]	72	56	1593
131	1	5197	2600	gi 153952	polymerase III polymerase subunit (dnaE) [Salmonella typhimurium] pir[A45915][A45915 DNA-directed DNA polymerase (EC 2.7.7.7) III ipha chain - Salmonella typhimurium]	72	53	2598
141	4	1040	1978	gi 1405446	transketolase [Bacillus subtilis]	72	54	939
149	8	2819	2835	gi 606234	lecY [Escherichia coli]	72	44	285
149	17	5472	5245	gi 1304472	DNA polymerase (unidentified phycodnavirus clone ORU)	72	55	228
154	1	1	210	gi 1205620	ferritin like protein [Haemophilus influenzae]	72	40	210
155	1	2207	1320	gi 391610	[farnesyl] diphosphate synthase [Bacillus stearothermophilus] pir[JX0257][JX0257 geranyltransferase (EC 2.5.1.10) - Bacillus stearothermophilus]	72	57	888
180	1	2	328	gi 433630	A180 [Saccharomyces cerevisiae]	72	62	327
184	3	1145	3553	gi 1205110	virulence associated protein homolog [Haemophilus influenzae]	72	49	2409
195	2	1923	1279	gi 1001730	hypothetical protein [Synecocystis sp.]	72	45	645
206	13	14646	15869	gi 1064607	ORTHINE AMINOTRANSFERASE [Bacillus subtilis]	72	50	1224
209	2	462	932	gi 1204666	hypothetical protein (GB:X73124_53) [Haemophilus influenzae]	72	60	471
215	2	764	522	gi 881513	insulin receptor homolog [Drosophila melanogaster] pir[S57245][S57245 insulin receptor homolog - fruit fly (Drosophila melanogaster) (SUB 46-2146)]	72	63	243
224	1	2	790	gi 949974	lactose repressor [Staphylococcus xylosum]	72	54	789
233	1	1526	765	gi 1408493	homologous to SwissProt:YIDA_ECOLI hypothetical protein [Bacillus subtilis]	72	52	762
240	1	220	1485	gi 537049	ORF_0470 [Escherichia coli]	72	52	1266
245	1	3	1340	gi 1204578	hypothetical protein (GB:U06949_1) [Haemophilus influenzae]	72	46	1338

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
259	2	2108	1245	gi 1340128	ORF1 [Staphylococcus aureus]	72	59	864
304	2	285	1094	gi 1205330	glutamine-binding periplasmic protein [Haemophilus influenzae]	72	52	810
307	10	5326	9039	gi 1070015	protein-dependent [Bacillus subtilis]	72	53	288
315	1	517	260	gi 143399	quinol oxidase [Bacillus subtilis]	72	55	258
316	11	9622	9308	gi 1204445	hypothetical protein (SP27857) [Haemophilus influenzae]	72	58	315
337	3	926	1609	gi 487433	citrate synthase II [Bacillus subtilis]	72	55	684
364	7	12538	10493	gi 1510643	ferrous iron transport protein B [Methanococcus jannaschii]	72	53	2046
409	2	340	1263	gi 1402944	orfR1 gene product [Bacillus subtilis]	72	49	924
441	3	2177	1590	gi 312379	highly conserved among subacteria [Clostridium acetobutylicum]	72	48	588
453	6	2654	2505	pir 600601 BXEA	pir S34312 S34312 hypothetical protein V - Clostridium acetobutylicum	72	70	150
460	1	2	625	gi 1016162	antibacterial protein 3 - Staphylococcus haemolyticus	72	51	624
463	1	3253	1828	gi 666014	ABC transporter subunit [Cyanophora paradoxa]	72	60	1626
480	4	3047	3466	gi 433992	ATP synthase subunit epsilon [Bacillus subtilis]	72	53	420
502	1	1086	586	gi 310859	ORF2 [Synecococcus sp.]	72	50	501
519	1	81	1184	gi 1303704	yrkE [Bacillus subtilis]	72	54	1104
559	1	3	746	gi 1107530	caud gene product [Campylobacter coli]	72	56	744
575	1	1142	573	gi 1303866	Yqps [Bacillus subtilis]	72	56	570
671	1	2	592	gi 1204697	protein-export membrane protein [Haemophilus influenzae]	72	44	591
679	2	295	1251	gi 563258	virulence-associated protein 2 [Dichelobacter nodosus]	72	52	957
687	2	295	957	gi 1146216	44% identical amino acids with the Escherichia coli enba ahpase; putative [Bacillus subtilis]	72	49	663
837	1	1	435	gi 1146183	putative [Bacillus subtilis]	72	54	435
868	1	150	788	gi 137842	unknown [Bacillus subtilis]	72	55	639
922	1	130	432	gi 1088269	unknown protein [Acetobacter vinelandii]	72	58	303
941	1	2	238	gi 153929	NADPH-sulfite reductase flavoprotein component [Salmonella typhimurium]	72	49	237
980	1	840	421	gi 853767	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [Bacillus subtilis]	72	59	420

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1209	1	383	213	gi 144735	neurotoxin type B [Clostridium botulinum]	72	44	171
1469	2	671	474	gi 1203458	hypothetical protein (G8:025562_47) [Haemophilus influenzae]	72	63	198
1956	1	727	365	gi 154409	hexosephosphate transport protein [Salmonella typhimurium] pir B41853 B41853 hexose phosphate transport system regulatory protein uhpB - Salmonella typhimurium	72	44	363
2101	1	3	401	gi 1303950	YqiY [Bacillus subtilis]	72	50	399
2503	1	569	399	gi 149713	formate dehydrogenase (Methanobacterium formicicum) pfr A42712 A42712 formate dehydrogenase (EC 1.2.1.2) - ethanobacterium formicicum	72	56	171
2967	1	3	155	gi 1212729	YqjZ [Bacillus subtilis]	72	46	153
3004	1	367	185	gi 665999	hypothetical protein [Bacillus subtilis]	72	55	183
3109	1	278	141	gi 413968	lpa-44d gene product [Bacillus subtilis]	72	45	138
3171	1	3	287	gi 1515938	glutamate synthase (ferredoxin) [Synecocystis sp.] pfr S46957 S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - Synecocystis sp.	72	52	285
3771	1	26	367	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	72	63	342
3951	1	1	222	gi 1500409	M. jannaschii predicted coding region M1519 [Methanococcus jannaschii]	72	38	222
4190	1	721	362	gi 39956	l1Glc [Bacillus subtilis]	72	57	360
4444	1	3	347	gi 1009166	respiratory nitrate reductase [Bacillus subtilis]	72	55	345
6	2	911	1200	gi 1537095	ornithine carbamoyltransferase [Kochuria coli]	71	54	270
11	15	11350	10859	gi 532309	25 kDa protein [Escherichia coli]	71	47	492
19	2	1248	2435	gi 1244574	D-alanine:D-alanine ligase [Enterococcus hirae]	71	52	1188
21	2	898	1488	gi 149629	anthranilate synthase component 2 [Leptospira biflexa] pfr C32840 C32840 anthranilate synthase (EC 4.1.3.27) component II Leptospira biflexa	71	45	591
34	1	1	567	gi 1303983	YqjP [Bacillus subtilis]	71	59	567
37	3	3192	2806	gi 1209681	glutamate-rich protein [Bacillus firmus]	71	50	387
38	18	12250	12462	gi 927645	arginyl endopeptidase [Porphyromonas gingivalis]	71	50	213
39	3	1246	4431	gi S09411 S094	spolIIE protein - Bacillus subtilis	71	49	3186
53	14	15770	14760	gi 142611	branched chain alpha-keto acid dehydrogenase E1-alpha [Bacillus subtilis]	71	58	1011
54	11	13461	12625	gi 143014	lont repressor [Bacillus subtilis]	71	46	837

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
57	7	7152	5860	gi 508175	ERIC domain of PTS-dependent Gac transport and phosphorylation <i>Escherichia coli</i>	71	48	1293
57	18	13897	14334	gi 1063247	high homology to flavohemoprotein (Hemoglobin-like protein) of <i>Alcaligenes eutrophus</i> and <i>Saccharomyces cerevisiae</i> ( <i>Bacillus subtilis</i> )	71	56	438
62	16	9831	10955	gi 1303926	YqjG ( <i>Bacillus subtilis</i> )	71	54	1125
70	12	8505	8966	gi 147198	phnE protein ( <i>Escherichia coli</i> )	71	38	462
86	5	2394	2089	gi 904205	hypothetical protein ( <i>Bacillus subtilis</i> )	71	51	306
96	7	7601	8269	gi 709991	hypothetical protein ( <i>Bacillus subtilis</i> )	71	49	669
100	6	4822	5931	gi 1060848	Opine dehydrogenase ( <i>Arthrobacter sp.</i> )	71	45	1110
103	1	1062	532	gi 143089	lep protein ( <i>Bacillus subtilis</i> )	71	41	531
109	18	15312	15695	gi 413985	lpa-61d gene product ( <i>Bacillus subtilis</i> )	71	57	384
113	1	630	316	gi 663254	probable protein kinase ( <i>Saccharomyces cerevisiae</i> )	71	57	315
114	5	6598	5603	gi 141156	membrane bound protein ( <i>Bacillus subtilis</i> )	71	40	996
133	2	3087	1723	gi 1303913	YqjX ( <i>Bacillus subtilis</i> )	71	53	1365
149	19	6335	5895	gi 529650	G40P ( <i>Bacteriophage SP21</i> )	71	51	441
154	5	3635	3087	gi 425488	repressor protein ( <i>Streptococcus sobrinus</i> )	71	47	549
164	11	11354	11689	gi 49318	ORF4 gene product ( <i>Bacillus subtilis</i> )	71	52	316
169	5	1936	2745	gi 1403403	unknown ( <i>Mycobacterium tuberculosis</i> )	71	56	810
193	2	272	1234	gi 1303788	YqjM ( <i>Bacillus subtilis</i> )	71	49	963
205	1	1743	895	gi 1215694	Ging ( <i>Mycoplasma pneumoniae</i> )	71	46	849
233	4	1849	2022	gi 433732	ORF1 ( <i>Campylobacter jejuni</i> )	71	50	174
237	7	4501	5169	gi 149384	HistE ( <i>Lactococcus lactis</i> )	71	54	669
272	4	2848	2273	gi 709993	hypothetical protein ( <i>Bacillus subtilis</i> )	71	48	576
274	2	618	1496	gi 141035	NAD(P)H:glutamate-transfer RNA reductase ( <i>Bacillus subtilis</i> ) pirA35252A35252 5-aminolevulinate synthase (EC 2.3.1.37) - <i>acillus subtilis</i>	71	53	879
276	5	3349	2720	gi 103562	ORF210 ( <i>Escherichia coli</i> )	71	50	630
287	1	136	660	gi 110634	20 kDa protein ( <i>Streptococcus gordonii</i> )	71	53	525
288	6	3322	2771	gi 1256625	putative ( <i>Bacillus subtilis</i> )	71	47	552

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
301	6	3492	2461	gi 467417	similar to lysine decarboxylase [Bacillus subtilis]	71	57	1032
306	4	6607	5222	gi 1256618	transport protein [Bacillus subtilis]	71	56	1386
307	2	1536	925	gi 632683	orfC (Mycoplama capricolum)	71	45	612
310	5	5793	5146	gi 348052	acetoin utilization protein [Bacillus subtilis]	71	51	648
322	1	2	1303	gi 1001819	hypothetical protein [Synechocystis sp.]	71	46	1302
333	4	4171	3995	gi 467473	unknown [Bacillus subtilis]	71	57	177
350	2	548	922	gi 551879	ORF 1 [Lactococcus lactis]	71	55	375
375	4	1860	3071	gi 467447	unknown [Bacillus subtilis]	71	57	1212
380	5	1560	2102	gi 142557	ATP synthase b subunit [Bacillus megaterium]	71	49	387
414	2	251	637	gi 580904	homologous to r. coli rnpA [Bacillus subtilis]	71	57	1020
424	1	335	1354	gi 581305	l-lactate dehydrogenase [Lactobacillus plantarum]	71	66	432
436	4	3701	3270	gi P0501 P0501	phosphoribosylanthranilate isomerase (EC 5.3.1.24) - Bacillus subtilis (fragment)	71	49	1278
482	1	3	1280	gi 410142	ORFX18 [Bacillus subtilis]	71	56	429
525	3	2272	1844	gi 143370	phosphoribosylpyrophosphate amidotransferase (Pur-F; EC 2.4.2.14) Bacillus subtilis	71	47	691
529	4	2739	2047	gi 606150	ORF 1309 [Escherichia coli]	71	53	948
563	1	22	969	gi 1237015	ORF4 [Bacillus subtilis]	71	47	252
581	1	506	255	gi 1301730	72503.2 [Caenorhabditis elegans]	71	55	156
612	2	1068	913	gi 153966	fibriane 2 [Salmonella typhimurium]	71	50	654
613	1	1	654	gi 466778	lysine specific perase [Escherichia coli]	71	52	621
618	1	1243	623	gi 1166238	poly(A) polymerase [Bacillus subtilis]	71	53	585
630	1	1170	586	gi 1486243	unknown [Bacillus subtilis]	71	51	486
691	1	1126	641	gi 289260	comE ORF1 [Bacillus subtilis]	71	47	275
694	2	149	427	gi 12971	NADH dehydrogenase subunit V (aa 1-605) [Gallus gallus] ir[S10197/S10197 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain - chicken mitochondrion (S2C1)]	71	53	609
715	2	169	777	gi 1303830	Yqft [Bacillus subtilis]	71	52	504
746	2	1473	970	gi 1377843	unknown [Bacillus subtilis]	71	52	504

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
748	1	1437	802	gi 1405459	YnsS [Bacillus subtilis]	71	49	636
753	1	1018	524	gi 1510389	M. jannaschii predicted coding region MJ0296 [Methanococcus jannaschii]	71	53	495
761	1	3	215	gi 475972	pentafunctional enzyme (Pneumocystis carinii)	71	47	213
783	1	1203	703	gi 536655	ORF YBR244w [Saccharomyces cerevisiae]	71	52	501
800	3	1292	987	gi 1204326	tRNA delta(2)-isopentenylpyrophosphate transferase [Haemophilus influenzae]	71	48	306
806	1	116	286	gi 1419075	cbiM gene product [Methanobacterium thermoautotrophicum]	71	50	171
931	1	973	488	gi 893358	PgsA [Bacillus subtilis]	71	56	486
1041	1	2	262	gi 1408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	71	45	261
1070	1	2	172	gi 709993	hypothetical protein [Bacillus subtilis]	71	46	171
1176	1	57	365	gi 151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas nevaloni] pir[A44756]A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	71	49	309
1181	1	366	184	gi 46971	epiP gene product [Staphylococcus epidermidis]	71	50	183
1281	1	3	290	gi 153016	ORF 419 protein [Staphylococcus aureus]	71	50	288
1348	1	436	229	gi 602683	orfC [Mycoplasma capricolum]	71	48	228
2002	1	756	379	gi 1008177	ORF YJL046w [Saccharomyces cerevisiae]	71	48	378
2119	1	2	217	gi 1046088	arginyl-tRNA synthetase [Mycoplasma genitalium]	71	50	216
2418	1	3	320	gi 1499771	M. jannaschii predicted coding region MJ0316 [Methanococcus jannaschii]	71	57	318
2961	3	2	187	gi 312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolycicus]	71	57	186
2999	2	67	306	gi 710020	nitrite reductase (nirB) [Bacillus subtilis]	71	43	240
3033	1	2	184	gi 1262335	YnsA [Bacillus subtilis]	71	57	183
3584	1	3	338	gi 401716	bata-isopropylmalate dehydrogenase [Neurospora crassa]	71	55	336
3715	2	743	399	gi 563952	gluconate permease [Bacillus licheniformis]	71	59	345
3785	1	770	387	gi 47382	acyl-CoA-dehydrogenase [Streptomyces purpurascens]	71	57	384
3875	1	541	272	gi 1001541	hypothetical protein [Synchocystis sp.]	71	38	270
4135	1	637	320	gi 142695	S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium	71	52	318
4249	1	63	239	gi 1203363	deoxyribose aldolase [Haemophilus influenzae]	71	63	177
4508	1	530	267	gi 1197667	vitellogenin [Anolis pulchellus]	71	46	264

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
6	3	1237	2721	gi 1321788	arginine ornithine antiporter [Clostridium perfringens]	70	54	1485
11	11	6572	7486	gi 216854	p47k [Pseudomonas chlororaphis]	70	41	915
12	1	2890	1481	gi 467330	replicative DNA helicase [Bacillus subtilis]	70	49	1410
15	1	1756	893	gi 451216	mannosephosphate isomerase [Streptococcus mutans]	70	46	864
15	2	1277	1050	gi 476092	unknown [Bacillus subtilis]	70	50	228
17	2	2132	1350	gi 145402	choline dehydrogenase [Escherichia coli]	70	52	783
21	1	2	925	gi 149516	anthranilate synthase alpha subunit [Lactococcus lactis] pir[S33124/S33124]	70	50	924
					anthranilate synthase (EC 4.1.3.27) alpha chain - actococcus lactis subsp. lactis			
25	7	5580	6251	gi 1389549	ORP3 [Bacillus subtilis]	70	52	672
33	6	6071	7423	gi 1303875	Yqha [Bacillus subtilis]	70	51	1353
36	2	959	1594	gi 500755	lactyl purine glycoylase [Mus musculus]	70	47	636
38	8	4901	5860	gi 1408507	pyrimidine nucleoside transport protein [Bacillus subtilis]	70	44	960
44	8	5312	5989	gi 1006620	hypothetical protein [Synecocystis sp.]	70	49	678
46	10	8950	10020	gi 1403126	cscD gene product [Alcaligenes eutrophus]	70	45	1071
52	2	2727	1900	gi 1486247	unknown [Bacillus subtilis]	70	53	828
52	6	4048	4656	gi 244501	esterase II:carboxylesterase [EC 3.1.1.1] [Pseudomonas fluorescens, optido, 218 aa]	70	50	609
56	8	8460	9962	gi 1339951	small subunit of MADII-dependent glutamate synthase [Plectonema boryanum]	70	51	1503
62	1	48	290	gi 142702	A competence protein 2 [Bacillus subtilis]	70	47	243
64	1	1080	541	gi 1204377	acetylcholinesterase biosynthesis protein [Haemophilus influenzae]	70	47	540
70	5	5139	3595	gi 1204034	2'-3'-cyclic-nucleotide 2'-phosphodiesterase [Haemophilus influenzae]	70	47	1545
91	4	7793	5466	gi 886471	methionine synthase [Catharanthus roseus]	70	56	2328
96	5	8754	7255	gi 1390561	alkaline phosphatase [EC 3.1.3.1] III precursor - Bacillus subtilis	70	54	1500
110	2	767	1300	gi 145294	adenine phosphoribosyl-transferase [Escherichia coli]	70	51	534
116	6	7026	7976	gi 143607	sporulation protein [Bacillus subtilis]	70	50	951
121	8	6401	6988	gi 1107528	ctg start [Campylobacter coli]	70	45	588
131	8	6842	7936	gi 1150454	proliferase PcpO [Lactobacillus delbrueckii]	70	48	1095

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
135	1	2	1489	gi 3111309	putative membrane-bound protein with four times repetition of ro-Sar-Ale at the N-terminus; function unknown [Alcaligenes utrophus]	70	49	1488
138	3	418	714	gi 504181	hypothetical protein [Bacillus subtilis]	70	46	297
164	8	9344	9874	gi 49315	ORF1 gene product [Bacillus subtilis]	70	47	531
164	16	15626	16618	gi 1205212	hypothetical protein (G8:D1048)_18 [Haemophilus influenzae]	70	50	993
205	2	2735	1803	gi 1215695	peptide transport system protein SapF homolog: SapF homolog [Mycoplasma pneumoniae]	70	47	933
209	3	910	1386	gi 1204665	hypothetical protein (G8:X73124_26) [Haemophilus influenzae]	70	48	477
246	3	340	756	gi 1215098	exciolase [Bacteriophage 154e]	70	46	417
263	7	7876	6749	gi 142540	aspartokinase II [Bacillus sp.]	70	51	1128
268	3	3212	4117	gi 1340128	ORF1 [Staphylococcus aureus]	70	50	906
302	6	3201	3827	gi 147782	ruvA protein (gtg start) [Escherichia coli]	70	46	627
302	10	5879	7051	gi 138530 C385	queline tRNA-ribosyltransferase (SC 2.4.2.29) - Escherichia coli	70	55	1173
313	1	2520	1416	gi 1205934	aminopeptidase a/1 [Haemophilus influenzae]	70	46	1107
355	2	379	669	gi 1070013	protein-dependent [Bacillus subtilis]	70	48	291
403	1	1255	629	gi 723147	GumF [Xanthomonas campestris]	70	33	627
444	10	8770	9273	gi 1204752	high affinity ribose transport protein [Haemophilus influenzae]	70	52	504
449	1	2	1243	gi 1619724	HgtE [Bacillus firmus]	70	44	1242
472	1	637	320	gi 727145	open reading frame; putative [Bacillus amyloquelaceus] p1r[B29091 B29091] hypothetical protein (bglA region) - Bacillus mytiloliquefaciens (fragment)	70	41	318
480	2	727	1608	gi 142560	ATP synthase gamma subunit [Bacillus megaterium]	70	44	882
524	1	2	307	gi 602292	KCH2 protein [Brassica napus]	70	45	306
525	1	823	413	gi 143372	phosphoribosyl glycineamide formyltransferase (Pur-N) [Bacillus subtilis]	70	52	411
565	4	1625	2552	gi 881434	ORF1 [Bacillus subtilis]	70	51	1074
607	4	829	1284	gi 151324	hypothetical protein (SP:P37002) [Methanococcus jannaschii]	70	50	456
633	1	1383	703	gi 431231	luciferase [Bacillus caldolyticus]	70	53	681
646	3	1683	1309	gi 467340	unknown [Bacillus subtilis]	70	49	375
663	1	830	417	gi 1303873	Y_032 [Bacillus subtilis]	70	40	414

TABLE 2

5. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
681	1	1488	781	gi 1001678	hypothetical protein [Synechocystis sp.]	70	53	708
708	1	2	448	sp P33940 YOJH_	HYPOTHETICAL 34.3 KD PROTEIN IN ECO-AUKB INTERGENIC REGION.	70	51	447
725	1	51	722	gi 1001644	hypothetical p-tein [Synechocystis sp.]	70	48	672
776	1	1171	787	gi 145165	putative [Escherichia coli]	70	47	585
834	1	250	783	gi 552971	NADH dehydrogenase (ndhF) [Vicia faba]	70	47	534
865	2	1585	1379	gi 1204636	ATP-dependent helicase [Haemophilus influenzae]	70	45	207
894	1	535	269	gi 1467364	DNA binding protein (probable) [Bacillus subtilis]	70	41	267
919	1	3	317	gi 1314847	CinA [Bacillus subtilis]	70	40	315
944	1	3	572	gi 709991	hypothetical protein [Bacillus subtilis]	70	44	570
988	2	772	605	gi 132441	ORF 3; putative [Bacillus subtilis]	70	50	168
1055	1	3	335	gi 529755	apeC [Streptococcus pyogenes]	70	37	333
1093	1	2	904	gi 1857754	ABC transporter [Bacillus subtilis]	70	49	903
1109	1	2	310	gi 1001827	hypothetical protein [Synechocystis sp.]	70	42	309
1220	1	468	235	pir S2316 S234	lepB protein - Staphylococcus epidermidis	70	40	234
1279	1	73	348	gi 153015	FtsH protein [Staphylococcus aureus]	70	47	276
1336	1	195	542	sp P31776 PDBA_	PENICILLIN-BINDING PROTEIN 1A (PBP-1A) (PENICILLIN-BINDING PROTEIN A).	70	50	348
1537	2	232	402	gi 1146181	putative [Bacillus subtilis]	70	50	171
1574	1	451	272	gi 219630	endothelin-A receptor [Homo sapiens]	70	47	180
1640	1	690	346	gi 11466243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative [Bacillus subtilis]	70	46	345
2504	1	2	286	gi 1495179	transmembrane protein [Lactococcus lactis]	70	51	285
3061	1	564	301	gi 508175	ERIC domain of prs-dependent GAT transport and phosphorylation Escherichia coli	70	44	264
3128	1	2	199	gi 1140096	unknown [Mycobacterium tuberculosis]	70	51	198
3218	1	3	488	gi 515938	glutamate synthase (ferredoxin) [Synechocystis sp.] pir S46957 S46957	70	50	486
3223	1	794	399	gi 1154891	glutamate synthase (ferredoxin) (EC 1.4.7.1) - ynechocystis sp.	70	52	396
3679	1	599	399	gi 529385	ATP binding protein [Photobacterium lewinsonii]	70	30	201
					chromosome condensation protein [Caenorhabditis elegans]	70	30	201

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3841	1	706	398	gi 1208965	hypothetical 23.3 kd protein [Escherichia coli]	70	47	309
3929	1	3	401	gi 149435	putative [Lactococcus lactis]	70	49	399
4044	1	595	374	gi 602031	similar to trimethylamine DII [Mycoplasma capricolus] pir[S49950]S49950 probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolus (S0C3) (fragment)	70	40	222
4129	1	558	280	gi 1339951	small subunit of NADH-dependent glutamate synthase [Plectonema boryanum]	70	49	279
4422	1	576	289	gi 296464	ATPase [Lactococcus lactis]	70	57	288
4647	1	361	200	gi 166412	NADH-glutamate synthase [Medicago sativa]	70	59	162
16	8	7571	9031	gi 1499620	M. jannaschii predicted coding region MJ0798 [Methanococcus jannaschii]	69	44	1461
16	9	9080	10033	gi 1353197	thioredoxin reductase [Eubacterium acidaminophilum]	69	54	954
30	1	1452	727	gi 1204910	hypothetical protein (GB:U14003_302) [Haemophilus influenzae]	69	52	726
38	4	1023	1298	gi 407773	devA gene product [Anabaena sp.]	69	41	276
44	9	5987	6595	gi 1205920	molybdate uptake system hydrophilic membrane-bound protein [Haemophilus influenzae]	69	45	609
62	15	9104	9475	gi 385178	unknown [Bacillus subtilis]	69	44	372
66	4	2402	2803	gi 1303893	YqhL [Bacillus subtilis]	69	51	402
67	15	14124	13627	gi 149647	ORF2 [Listeria monocytogenes]	69	37	498
67	17	14033	14382	gi 305002	OMP_F356 [Escherichia coli]	69	49	330
67	19	15130	15807	gi 1109684	ProV [Bacillus subtilis]	69	45	678
78	3	1447	2124	gi 1256633	putative [Bacillus subtilis]	69	53	678
78	4	4513	3725	gi 1303958	YqjG [Bacillus subtilis]	69	32	789
85	4	4521	4213	pir E29326 E293	hypothetical protein (pur operon) - Bacillus subtilis	69	32	309
86	6	3253	2654	gi 173332	OrfC [Bacillus subtilis]	69	50	600
95	1	96	710	gi 766468	14kD antigen, sperm tail membrane antigen-putative sucrose-specific phosphotransferase enzyme II homolog [rice, testis, Peptid. Partial, 72 aa]	69	43	615
100	7	6023	7426	gi 1205355	Na+/H+ antiporter [Haemophilus influenzae]	69	39	1404
102	2	2678	1650	gi 561690	islatoglycoproteinase [Pasteurella hemolytica]	69	47	1029
103	8	12241	8537	gi 1009366	respiratory nitrate reductase [Bacillus subtilis]	69	54	3705
103	11	14987	12552	gi 710020	nitrite reductase (nirB) [Bacillus subtilis]	69	51	2436

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
112	11	8708	10168	gi 15411	hexosephosphate transport protein (Salmonella typhimurium) p1r p1833 p1833 hexose phosphate transport system protein uhpT - salmonella typhimurium	69	51	1461
112	16	16644	17414	gi 1204835	pyruvate formate-lyase activating enzyme (Haemophilus influenzae)	69	50	771
113	2	33	953	gi 290509	loj07 (Escherichia coli)	69	43	921
114	2	1537	1058	p1r A4271 A427	reticulocyte-binding protein 1 - Plasmodium vivax	69	39	480
121	6	4309	5310	gi 1154633	MrpF (Bacillus subtilis)	69	53	1002
125	2	267	854	gi 413931	lpa-7d gene product (Bacillus subtilis)	69	43	588
149	127	10666	110400	p1r S28089 S280	hypothetical protein A - yeast (Zygosaccharomyces bisporus) plasmid pS0	69	39	267
161	1	1598	813	gi 1205538	hypothetical protein (GB:U14003_302) (Haemophilus influenzae)	69	47	786
165	4	2222	4633	gi 140054	phenylalanyl-tRNA synthetase beta subunit (AA 1-804) (Bacillus subtilis)	69	52	2412
169	3	1210	1761	gi 1296031	elongation factor Ts (Spirulina platensis)	69	45	552
175	12	8686	8339	gi 732682	PIA5 protein (Escherichia coli)	69	69	348
190	2	484	1671	ep p17731 H158	HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (INDAZOLE ACETOL-PHOSPHATE TRANSAMINASE)	69	48	1188
206	1	5551	2777	gi 41750	hadr protein (AA 1-1033) (Escherichia coli)	69	49	2775
206	4	6038	5796	gi 1256135	Ybnp (Bacillus subtilis)	69	46	243
249	1	636	319	gi 1405436	YnuP (Bacillus subtilis)	69	50	118
302	8	4820	5776	gi 1001768	hypothetical protein (Synecocystis sp.)	69	46	957
324	2	7384	3893	gi 1256798	pyruvate carboxylase (Rhitobium etli)	69	53	3492
351	3	2098	1808	gi 1491664	T04H1.4 (Caenorhabditis elegans)	69	30	291
369	3	2075	2305	gi 336458	ORP (Balaenoptera acutorostrata)	69	61	231
392	3	1999	2424	gi 558015	ORP (Bacillus subtilis)	69	45	426
410	1	87	779	gi 355611	phosphoglyceromutase (Zymomonas mobilis)	69	58	693
421	1	2085	1129	gi 1276985	arginase (Bacillus caldovelox)	69	54	957
444	8	6713	7741	gi 1221782	purine synthesis repressor (Haemophilus influenzae)	69	40	1029
453	1	828	415	gi 1122758	unknown (Bacillus subtilis)	69	57	414
469	2	3286	2246	gi 1458228	mutY homolog (Homo sapiens)	69	44	1041

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match name	% sim	% ident	length (nt)
509	3	1730	1371	gi 49224	JURF 4 [Synecoccus sp.]	69	39	360
520	5	3023	2823	gi 726427	similar to D. melanogaster H57101-2 protein (PIR:S34154) Caenorhabditis elegans	69	39	201
531	1	26	760	gi 509672	repressor protein (Bacteriophage Tuc2009)	69	33	735
589	1	107	253	gi 169101	17.9 kDa heat shock protein (hsp17.9) [Pisum sativum]	69	52	147
594	2	597	1191	gi 142783	DNA photolyase [Bacillus firmus]	69	48	795
604	4	2476	2114	gi 413930	Ips-6d gene product [Bacillus subtilis]	69	45	363
607	1	2	313	gi 1236103	W08D2.3 [Caenorhabditis elegans]	69	47	312
607	2	590	312	gi 1536715	ORF YBR275c [Saccharomyces cerevisiae]	69	39	279
714	1	864	433	gi 467327	unknown [Bacillus subtilis]	69	44	432
759	1	3	338	gi 1009367	Respiratory nitrate reductase [Bacillus subtilis]	69	50	336
761	2	392	586	gi 3508	Leucyl-tRNA synthetase (cytoplasmic) [Saccharomyces cerevisiae]	69	46	195
802	1	72	1013	gi 143044	ferrochelatase [Bacillus subtilis]	69	55	942
816	1	2573	1368	gi 1510268	restriction modification system 5 subunit [Methanococcus jannaschii]	69	45	1206
838	2	133	387	gi 1255371	coded for by C. elegans cDNA YK349.5; coded for by C. elegans cDNA YK349.1; Similar to guanylate kinase [Caenorhabditis elegans]	69	46	255
851	2	745	1005	gi 288998	scdA gene product [Antilimonium sp.]	69	39	261
867	1	535	269	gi 1070014	protein-dependent [Bacillus subtilis]	69	47	267
995	1	954	478	gi 1205569	transcription elongation factor [Haemophilus influenzae]	69	53	477
999	1	1009	506	gi 899251	predicted trichorax protein [Drosophila virilis]	69	21	504
1127	1	1315	659	gi 1205434	H. influenzae predicted coding region H1191 [Haemophilus influenzae]	69	56	657
1138	1	248	460	gi 1510646	M. jannaschii predicted coding region MJ0568 [Methanococcus jannaschii]	69	48	213
2928	1	3	401	gi 20503	glutamate permease [Escherichia coli]	69	41	399
3090	1	444	223	gi 1204987	DNA polymerase III, alpha chain [Haemophilus influenzae]	69	36	222
3817	1	2	400	gi 1483199	peptide-synthetase [Mycobacterium mageritense]	69	45	399
3833	1	667	335	gi 1524193	unknown [Mycobacterium tuberculosis]	69	46	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
4079	1	747	400	gi 546918	orf13' of comK [Bacillus subtilis, 826, Peptide Partial, 140 aa] pir S43612 S43612 hypothetical protein Y - Bacillus subtilis sp P03098 HXD_BACSD HYPOTHETICAL PROTEIN IN COMK 3'REGION (ORFY) FRAGMENT)	69	64	348
4115	2	215	400	gi 517205	167 kDa Myosin-crossreactive streptococcal antigen [Streptococcus pyogenes]	69	59	186
4139	1	1	333	gi 1208451	hypothetical protein [Synchocystis sp.]	69	36	333
4258	1	457	230	gi 496158	restriction-modification enzyme subunit M1 [Mycoplasma pneumoniae] pir S49395 S49395 Mdh1 protein - Mycoplasma pneumoniae [SOC3]	69	43	228
4317	1	90	374	gi 413967	ipa-43d gene product [Bacillus subtilis]	69	44	285
4465	1	3	293	gi 396296	similar to phosphotransferase system enzyme II [Escherichia coli] sp P32672 PNC_ECOLI PTS SYSTEM, FRUCTOSE-1,6-BP-2 ITC COMPONENT PHOSPHOTRANSFERASE ENZYME II, C COMPONENT)	69	49	291
3	1	2102	1193	gi 1109685	ProW [Bacillus subtilis]	68	46	1110
15	4	2592	2074	gi 807973	unknown [Saccharomyces cerevisiae]	68	45	519
31	8	6328	8772	gi 290642	ATPase [Enterococcus hirae]	68	48	2445
40	2	1115	750	gi 606342	ORF_0522, reading frame open far upstream of start; possible rameshift, linking to previous ORF [Escherichia coli]	68	55	166
46	9	8886	8415	gi 355276	aldehyde dehydrogenase [Vibrio cholerae]	68	44	1530
48	3	3643	3404	gi 285608	241k polypeptide [Apple stem grooving virus]	68	47	240
48	4	3536	4132	gi 1045937	M. genitalium predicted coding region M246 [Mycoplasma genitalium]	68	39	597
53	10	11671	10685	gi 3303952	VqJA [Bacillus subtilis]	68	46	987
70	9	7346	8155	gi 147198	phnE protein [Escherichia coli]	68	40	810
89	4	1899	2966	gi 145173	35 kDa protein [Escherichia coli]	68	43	1068
108	1	2187	1150	gi 38722	precursor (aa -20 to 381) [Acinetobacter calcoaceticus] ir A29277 A29277 aldose 1-epimerase (EC 5.1.3.3) - Acinetobacter calcoaceticus	68	57	1038
112	5	2666	3622	gi 353724	MalC [Streptococcus pneumoniae]	68	55	957
116	7	7865	8638	gi 343608	sporulation protein [Bacillus subtilis]	68	48	774
118	3	2484	3698	gi 3303805	YgeR [Bacillus subtilis]	68	46	1215
120	2	1424	1594	sp P38038 CYBQ_	SULFITE REDUCTASE (NADPH) FLAVOPROTEIN ALPHA-COMPONENT (EC 1.8.1.3) (SIR- PPI)	68	45	171
129	1	1	1011	gi 396307	argininosuccinate lyase [Escherichia coli]	68	50	1011

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
132	3	1867	gi 116267	ORF2 [Bacillus megaterium]	68	48	873
134	2	888	gi 147545	DNA recombinase [Escherichia coli]	68	50	165
141	2	372	gi 872116	atc (strains inducible protein) [Glycine max]	68	36	243
149	7	2454	gi 145774	hsp70 protein (dnaK gene) [Escherichia coli]	68	48	195
155	2	1776	gi 216583	ORF1 [Escherichia coli]	68	38	243
158	3	1826	sp P33940 YQJH_	HYPOPHOSPHATE 54.3 KD PROTEIN IN ECO-NLKB INTERGENIC REGION.	68	51	1464
169	6	2749	gi 1403402	unknown [Mycobacterium tuberculosis]	68	46	570
175	10	9158	gi 1072395	phaA gene product [Rhizobium meliloti]	68	51	1794
188	7	4184	gi 1173843	3-ketoacyl-ACP synthase II [Vibrio Harveyi]	68	48	975
189	3	907	gi 467383	DNA binding protein (probable) [Bacillus subtilis]	68	48	975
206	5	7683	gi 1256138	YbbI [Bacillus subtilis]	68	48	1752
206	8	10425	gi 452687	pyruvate decarboxylase [Saccharomyces cerevisiae]	68	39	228
212	8	3421	gi 1369941	lcl gene product [Bacteriophage B1]	68	45	1026
214	8	5457	gi 1420467	ORF YOR196c [Saccharomyces cerevisiae]	68	46	582
237	4	2507	gi 149381	HlsR [Lactococcus lactis]	68	47	999
243	5	5340	gi 1235684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	68	42	162
262	1	3	gi 150974	4-oxalocrotonate tautomerase [Pseudomonas putida]	68	49	867
262	2	1984	gi 1147744	PSR [Enterococcus hirae]	68	50	564
276	6	3702	sp P30750 ABC_E	ATP-BINDING PROTEIN ABC (FRAGMENT)	68	53	621
306	6	6345	gi 1256617	adenine phosphoribosyltransferase [Bacillus subtilis]	68	45	750
333	3	4599	gi 467473	unknown [Bacillus subtilis]	68	45	180
365	6	5017	gi 1130643	T22B3.3 [Caenorhabditis elegans]	68	51	1098
376	2	549	gi 1277026	DAPA aminotransferase [Bacillus subtilis]	68	47	870
405	1	1741	gi 1103917	YqjB [Bacillus subtilis]	68	44	315
406	2	853	gi 1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	68	48	168
426	6	3558	gi 624632	GlcL [Escherichia coli]	68	43	222
438	1	108	gi 146923	nitrogenase reductase [Escherichia coli]	68	43	222

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
443	1	476	240	gi 535810	hippuricase [Campylobacter jejuni]	68	42	237
443	2	518	1015	gi 1204742	H. influenzae predicted coding region H10491 [Haemophilus influenzae]	68	48	498
443	5	4447	3779	gi fn660	deoxyribose-phosphate aldolase [Bacillus subtilis] pir 549455 549455 deoxyribose-phosphate aldolase [EC 4.1.2.4] - acillus subtilis	68	55	669
476	2	240	1184	gi 571345	unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp P45860 YME_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA INTERGENIC REGION.	68	45	945
486	2	1876	1046	gi 147328	transport protein [Escherichia coli]	68	41	831
517	3	1764	2084	gi 521809	orf2 [Bacteriophage A2]	68	64	321
572	1	2	571	sp P19237 Y0SL	HYPOTHETICAL 6.8 KD PROTEIN IN NRDC-TK INTERGENIC REGION.	68	47	570
646	1	914	459	gi 413982	ipa-58r gene product [Bacillus subtilis]	68	52	456
659	3	1668	1901	gi 1107541	C3309.8 [Caenorhabditis elegans]	68	36	234
864	5	1510	1716	gi 145774	hap70 protein (dnaK gene) [Escherichia coli]	68	48	207
920	1	860	432	gi 1510416	hypothetical protein (SP:P31466) [Methanococcus jannaschii]	68	54	429
952	1	1096	611	gi C33456	reductase [Leishmania major]	68	46	486
970	1	91	402	gi 1354775	pfoS/R [Treponema pallidum]	68	46	312
1028	1	1064	534	gi 151017	diaminopimelate decarboxylase [Bacillus subtilis]	68	47	531
1029	1	428	216	gi 1135714	Plasmidium falciptarum mRNA for asparagine-rich antigen (clone 17C1) [Plasmidium falciptarum]	68	31	213
1058	1	692	348	gi 181649	epic gene product [Staphylococcus epidermidis]	68	46	345
1096	2	665	465	gi 143434	Rho Factor [Bacillus subtilis]	68	43	201
1308	1	2	694	gi 1169939	group B oligopeptidase.PapB [Streptococcus agalactiae]	68	50	693
1679	1	2	238	gi 17205	67 kDa Myosin-crossreactive streptococcal antigen [Streptococcus yogenes]	68	53	237
2039	1	3	383	gi 153898	transport protein [Salmonella typhimurium]	68	51	381
2077	1	3	326	pir C33496 C334	hisC homolog - Bacillus subtilis	68	47	324
2112	1	613	374	gi 64884	lamin B1 [Xenopus laevis]	68	50	240
2273	1	793	398	gi 581648	epiB gene product [Staphylococcus epidermidis]	68	45	396
2848	1	2	385	gi 216869	branched-chain amino acid transport carrier [Pseudomonas aeruginosa] pir A38534 A38534 branched-chain amino acid transport protein braz Pseudomonas aeruginosa	68	41	384

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
2955	1	768	400	gi 504179	hypothetical protein [Bacillus subtilis]	68	49	369
2981	1	572	288	gi 508979	GTP-binding protein [Bacillus subtilis]	68	48	285
3014	1	584	294	gi 1524394	ORF-2 upstream of gbaB operon [Bacillus subtilis]	68	45	291
3082	1	336	169	gi 1204696	fructose-permease 18C component [Haemophilus influenzae]	68	53	168
3108	1	103	258	gi 217855	heat-shock protein [Arabidopsis thaliana]	68	48	156
3639	1	919	461	gi 1510490	nitrate transport permease protein [Methanococcus jannaschii]	68	47	459
3657	1	1	330	gi 155369	PTS enzyme-II fructose [Xanthomonas campestris]	68	48	330
3823	1	780	391	gi 603768	HutI protein, imidazolone-5-propionate hydrolase [Bacillus subtilis]	68	54	390
3982	1	2	277	gi 149435	putative [Lactococcus lactis]	68	47	276
4051	1	1	342	gi 450688	hadM gene of Ecoprr1 gene product [Escherichia coli] pir S38437 S38437 hadM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	48	342
4089	1	12	209	gi 135378	heavy-metal transporting P-type ATPase [Proteus mirabilis]	68	47	198
4143	1	47	187	gi 603769	HutU protein, urecanase [Bacillus subtilis]	68	55	141
4148	1	2	352	gi 450688	hadM gene of Ecoprr1 gene product [Escherichia coli] pir S38437 S38437 hadM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	51	351
4173	1	2	382	gi 1041097	Pyruvate Kinase [Bacillus psychrophilus]	68	48	381
4182	1	498	750	gi 413968	ipa-44d gene product [Bacillus subtilis]	68	50	249
4362	2	148	318	gi 450688	hadM gene of Ecoprr1 gene product [Escherichia coli] pir S38437 S38437 hadM protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	68	44	171
5	11	9493	8300	gi 143727	putative [Bacillus subtilis]	67	46	1194
31	11	10318	9833	gi 116746	D-lactate dehydrogenase [Lactobacillus plantarum]	67	41	486
32	3	1560	3155	gi 1098557	renal sodium/dicarboxylate cotransporter [Homo sapiens]	67	46	1596
32	5	4945	4145	gi 1510720	prephenate dehydratase [Methanococcus jannaschii]	67	51	801
36	5	5350	4268	gi 146216	45% identity with the product of the ORF6 gene from the <i>Erwinia herbicola</i> carotenoid biosynthesis cluster; putative [Bacillus subtilis]	67	58	1083
44	7	4492	5304	gi 1006621	hypothetical protein [Synecocystis sp.]	67	43	813

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
56	7	3943	8481	gi 356131	glutamate synthase large subunit precursor [Azospirillum brasilense] p1r186602 186602 glutamate synthase (NADPH) (EC 1.4.1.13) alpha chain - Azospirillum brasilense	67	52	4539
56	12	11923	14678	gi 1000453	TrpR [Bacillus subtilis]	67	48	756
62	8	5092	4757	gi 1113949	orf3 [Bacillus, C-125, alkali-sensitive mutant 18224, Peptide Mutant, 112 aa]	67	45	336
62	10	17570	6338	gi 654655	Na/H antiporter system [Bacillus alcalophilus]	67	49	1233
99	3	2119	3321	gi 1204349	hypothetical protein (GB:D90212.3) [Haemophilus influenzae]	67	50	1203
102	9	5695	7176	gi 149432	putative [Lactococcus lactis]	67	51	1482
103	11	14549	14049	gi 1408497	[LPSD gene product [Bacillus subtilis]	67	48	501
109	15	14821	13982	gi 413976	lipa-52r gene product [Bacillus subtilis]	67	49	840
109	17	14811	15194	gi 413983	lipa-59d gene product [Bacillus subtilis]	67	29	384
121	4	1713	2153	gi 1262335	YmaA [Bacillus subtilis]	67	54	441
122	1	1	1149	gi 143047	ORF8 [Bacillus subtilis]	67	35	1149
124	5	4060	3518	gi 556885	unknown [Bacillus subtilis]	67	47	543
131	2	4584	3589	gi 1046081	hypothetical protein (GB:D26185.10) [Mycoplasma genitalium]	67	30	996
140	3	2899	2297	gi 146549	kdpC [Escherichia coli]	67	45	603
142	4	5409	4198	gi 1212775	GTP cyclohydrolase II [Bacillus amyloliquefaciens]	67	55	1212
147	5	2913	2374	gi 1303709	YrkJ [Bacillus subtilis]	67	44	540
152	8	6341	6673	gi 1377841	unknown [Bacillus subtilis]	67	48	333
161	4	2720	3763	gi 496319	SphX [Synecococcus sp.]	67	47	1044
163	6	1989	3428	gi 595681	2-oxoglutarate/malate translocator [Spinacia oleracea]	67	53	276
193	3	1351	1626	gi 1511101	shikimate 5-dehydrogenase [Methanococcus jannaschii]	67	48	1263
200	2	917	2179	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	67	47	357
206	10	12445	112801	ep P37347 YEC0_	HYPOHETICAL 21.8 KD PROTEIN IN ASPS 5'REGION.	67	47	357
206	11	13047	14432	gi 732813	branched-chain amino acid carrier [Lactobacillus delbrueckii]	67	46	1386
208	2	1321	809	gi 1033037	100 kDa heat shock protein (Hsp100) [Leishmania major]	67	36	513
238	3	1039	2052	gi 609542	CbrB protein [Erwinia chrysanthemi]	67	42	1014

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
246	2	176	367	gi 215098	excisionase [Bacteriophage 154a]	67	37	192
276	2	2260	1412	gi 303560	ORP271 [Escherichia coli]	67	50	849
297	6	2223	3056	gi 142784	CtsA protein [Bacillus firmus]	67	46	834
307	7	5220	4186	gi 1070013	protein-dependent [Bacillus subtilis]	67	43	1035
316	1	36	1028	gi 1161061	diacylglycerol kinase [Methylobacterium extorquens]	67	52	993
324	3	5650	5030	gi 1469784	putative cell division protein ftsW [Enterococcus hirae]	67	49	621
336	1	524	264	gi 173122	urea amidolyase [Saccharomyces cerevisiae]	67	45	261
360	1	104	1194	gi 130053	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) [HISTIDINE--TRNA LIGASE] (HISMG)	67	47	1287
364	3	4890	3592	gi 151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonil] pir A44756 A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	67	46	1299
365	3	2940	2113	gi 1296823	orf2 gene product [Lactobacillus helveticus]	67	47	828
367	2	325	918	gi 1039479	ORF1 [Lactococcus lactis]	67	47	594
395	3	666	1271	gi 1204516	hypothetical protein (GB:00014_4) [Haemophilus influenzae]	67	55	606
415	1	1800	901	gi 382579	CG Site No. 29739 [Escherichia coli]	67	46	900
419	1	1799	903	gi 520752	putative [Bacillus subtilis]	67	48	897
474	1	2	796	gi 446906	argininosuccinate synthetase [Streptomyces clavuligerus] pir S57659 S57659 argininosuccinate synthase (EC 6.3.4.5) - Streptomyces clavuligerus	67	49	791
485	2	1921	2226	gi 143434	Rho factor [Bacillus subtilis]	67	43	306
596	1	1728	865	gi 1303853	YqgF [Bacillus subtilis]	67	47	864
700	1	433	218	gi 1204628	hypothetical protein (SP:E21498) [Haemophilus influenzae]	67	47	216
806	2	249	647	gi 677947	AppC [Bacillus subtilis]	67	51	399
828	2	340	900	gi 777761	IrrA [Synecoccus sp.]	67	37	561
833	1	1407	916	gi 142996	regulatory protein [Bacillus subtilis]	67	41	492
856	1	1555	779	gi 780224	2K970.2 [Caenorhabditis elegans]	67	38	777
888	1	1614	850	gi 437315	TTG start codon [Bacillus licheniformis]	67	40	765
1034	1	1190	597	gi 1205113	hypothetical protein (GB:19201_15) [Haemophilus influenzae]	67	45	594
1062	1	636	319	gi 1303850	YqgC [Bacillus subtilis]	67	41	318
1067	1	918	460	pir A32950 A329	probable reductase protein - Leishmania major	67	54	459

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1358	1	3	293	gi 1001369	hypothetical protein (Synecocystis sp.)	67	44	291
2181	1	3	302	gi 1510416	hypothetical protein (SP:P31666) (Methanococcus jannaschii)	67	48	300
3000	1	1	507	gi 517205	67 kDa Myosin-crossreactive streptococcal antigen (Streptococcus yogeneae)	67	56	507
3066	1	464	234	gi 308861	GTC start codon (Lactococcus lactis)	67	46	231
3087	1	454	251	gi 1205366	oligopeptide transport ATP-binding protein (Haemophilus influenzae)	67	44	206
3101	1	2	256	gi 1531541	uroporphyrinogen III methyltransferase (Zea mays)	67	55	255
3598	1	728	393	gi 151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir A44756 (EC 1.1.1.88) Pseudomonas sp. hydroxymethylglutaryl-CoA reductase	67	56	336
3765	2	584	366	gi 557489	menD (Bacillus subtilis)	67	45	219
3788	1	638	398	pir S52915 S529	nitrate reductase alpha chain - Bacillus subtilis (fragment)	67	45	261
3883	1	2	265	gi 704397	cystathionine beta-lyase (Arabidopsis thaliana)	67	46	264
3926	1	2	340	gi 1483199	peptide-synthetase (Amycolatopsis mediterranea)	67	44	339
4417	1	82	396	gi 1205337	ribonucleotide transport ATP-binding protein (Haemophilus influenzae)	67	46	315
2	3	3075	3989	gi 533348	CodV (Bacillus subtilis)	66	42	915
15	6	2273	2542	gi 46491	SacB (Synecococcus PCC7942)	66	37	270
31	9	8059	7826	gi 292046	mucin (Homo sapiens)	66	44	234
31	10	9014	9258	gi 1204545	mercury scavenger protein (Haemophilus influenzae)	66	48	225
32	6	6347	5253	gi 998342	inducible nitric oxide synthase (Callus gallus)	66	47	1095
44	13	8856	10124	gi 1510751	molybdenum cofactor biosynthesis moeA protein (Methanococcus jannaschii)	66	46	1269
48	2	1276	2868	gi 150209	ORF 1 (Mycoplasma mycoides)	66	40	1593
58	8	7178	8428	gi 665999	hypothetical protein (Bacillus subtilis)	66	47	1251
62	7	5143	4370	gi 1072398	phaD gene product (Rhizobium mallei)	66	40	774
70	14	11693	10998	gi 809660	deoxyribose-phosphate aldolase (Bacillus subtilis) pir S49455 S49455 deoxyribose-phosphate aldolase (EC 4.1.2.4) - acillus subtilis	66	55	696
76	1	1	1305	gi 142440	ATP-dependent nuclease (Bacillus subtilis)	66	42	1305
91	6	9236	8205	gi 704397	cystathionine beta-lyase (Arabidopsis thaliana)	66	43	1032
102	5	3810	3265	gi 1204323	hypothetical protein (SP:P31805) (Haemophilus influenzae)	66	41	546

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	seqch gene name	% sim	% ident	length (nt)
103	4	3418	2732	gi 971344	nitrate reductase gamma subunit [Bacillus subtilis] ap P42177 NARI_BACSU NITRATE REDUCTASE GAMMA CHAIN (EC 1.7.99.4) - gi 1009169 Respiratory nitrate reductase [Bacillus subtilis] (Sub -160)	66	48	687
109	6	4243	4674	gi 170886	glucosamine-6-phosphate deaminase [Candida albicans] pir A6652 A6652 glucosamine-6-phosphate isomerase (EC 5.3.1.10) - east (Candida albicans)	66	45	432
112	17	17491	17712	gi 1323179	ORF YCR111W [Saccharomyces cerevisiae]	66	32	222
116	2	4667	2637	gi 1491813	gamma-glutamyltranspeptidase [Bacillus subtilis]	66	43	2031
130	5	3189	2989	gi 1146224	putative [Bacillus subtilis]	66	30	201
172	5	3264	3662	gi 755152	highly hydrophobic integral membrane protein [Bacillus subtilis] ap P42953 TAGC_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGO.	66	41	399
174	5	4592	3723	gi 1146241	pantothenate synthetase [Bacillus subtilis]	66	49	870
175	4	3209	2880	gi 642655	unknown [Rhizobium melliottii]	66	29	330
175	11	8743	7994	gi 854655	Na/H antiporter system [Bacillus alcalophilus]	66	43	750
190	5	7079	5727	gi 451072	di-tripeptide transporter [Lactococcus lactis]	66	40	1351
195	15	13919	13713	gi 1322411	unknown [Mycobacterium tuberculosis]	66	42	207
217	3	2822	2595	gi 1143542	alternative stop codon [Rattus norvegicus]	66	36	228
233	9	7133	6135	gi 1458327	F08P3.4 gene product [Caenorhabditis elegans]	66	47	999
238	1	41	1041	gi 809541	CurA protein [Brwinia chrysanthum]	66	42	999
241	1	2102	1053	gi 153067	peptidoglycan hydrolase [Staphylococcus aureus]	66	53	1050
261	1	1178	648	gi 1510859	M. jannaschii predicted coding region MJ0790 [Methanococcus jannaschii]	66	40	531
263	3	3731	2973	gi 1205865	tetrahydrodipicolinate N-succinyltransferase [Haemophilus influenzae]	66	47	759
272	8	6548	5484	gi 882101	high affinity nickel transporter [Alcaligenes eutrophus] ap P23516 HOUX_ALCEU HIGH-AFFINITY NICKEL TRANSPORT PROTEIN.	66	44	1065
276	3	2805	2104	gi 1208965	hypothetical 23.3 kd protein [Escherichia coli]	66	47	702
278	2	2830	1784	gi 1488662	phosphatase-associated protein [Bacillus subtilis]	66	48	1047
278	3	3830	2952	gi 303560	ORF271 [Escherichia coli]	66	45	879
279	2	3894	2218	gi 1185289	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1- carboxylate synthase [Bacillus subtilis]	66	48	1677
288	4	2535	2275	gi 1256625	putative [Bacillus subtilis]	66	42	261
292	2	1133	942	gi 1511604	M. jannaschii predicted coding region MJ165 [Methanococcus jannaschii]	66	30	192

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
294	1	1116	559	gi 216314	esterase [Bacillus stearothermophilus]	66	45	558
297	4	2913	1978	gi 994794	cytochrome a assembly facto [Bacillus subtilis] sp P24009 COX2_BACSU PROBABLE CYTOCHROME C OXIDASE ASSEMBLY FACTOR.	66	45	936
316	4	2053	2682	gi 1107839	lactate lyase [Pseudomonas aeruginosa]	66	40	630
338	4	2460	2302	gi 520750	biotin synthetase [Bacillus sphaericus]	66	59	159
339	1	1214	735	gi 467468	7, 8-dihydro-5-hydroxymethylpterin-pyrophosphokinase [Bacillus subtilis]	66	52	480
363	1	3	863	gi 581649	epIC gene product [Staphylococcus epidermidis]	66	47	861
366	2	232	483	gi 1103505	unknown [Schizosaccharomyces pombe]	66	53	252
367	4	2468	1845	sp P20692 TYRA_	PREPHENATE DEHYDROGENASE (EC 1.3.1.12) (PDH)	66	50	624
372	3	2150	1599	gi 467416	unknown [Bacillus subtilis]	66	38	552
378	1	212	1009	gi 147309	purine nucleoside phosphorylase [Escherichia coli]	66	50	798
401	1	1	462	gi 388263	p-aminobenzoic acid synthase [Streptomyces griseus] p JN0531 JN0531 p-aminobenzoic acid synthase - Streptomyces riseus	66	46	462
404	7	4826	5254	gi 606744	cytidine deaminase [Bacillus subtilis]	66	51	429
411	2	1738	1103	gi 1460081	unknown [Mycobacterium tuberculosis]	66	44	636
420	1	2	541	gi 1046024	Na+ ATPase subunit J [Mycoplasma genitalium]	66	49	540
431	1	1	858	gi 1500008	M. jannaschii predicted coding region MJ1154 [Methanococcus jannaschii]	66	50	858
443	7	5679	5299	gi 852076	HrgA [Bacillus subtilis]	66	46	381
444	3	3405	2413	gi 153047	lysostaphin (lto start codon) [Staphylococcus simulans] p J25483 J25483 lysostaphin precursor - Staphylococcus simulans sp P10547 LSTP_STASI LYOSTAPHIN PRECURSOR (EC 3.5.1.-)	66	51	993
561	1	956	480	gi 1204905	DNA-3-methyladenine glycosylase I [Haemophilus influenzae]	66	45	477
562	3	1066	1383	gi 1046082	M. genitalium predicted coding region MG372 [Mycoplasma genitalium]	66	52	318
576	1	11	724	gi 305014	ORF_0234 [Escherichia coli]	66	43	714
577	3	1190	903	gi 1001353	hypothetical protein [Synachocystis sp.]	66	52	288
584	1	2	331	ep P24204 VEBA_	HYPOHETICAL 46.7 KD PROTEIN IN MSBB-RUVB INTERGENIC REGION (ORFU)	66	48	310
592	1	1410	706	gi 928839	ORF266: putative [Lactococcus lactis phase BK5-F]	66	51	705
601	1	1413	720	gi 1488695	novel antigen; orf-2 [Staphylococcus aureus]	66	55	714

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
619	3	468	845	gi1746573	similar to M. musculus transport system membrane protein, Nmap PIR:A0739 and S. cerevisiae Shf1 protein (PIR:A4536) <i>Caenorhabditis elegans</i>	66	45	378
706	2	561	355	gi1804808	unknown protein [ <i>Rattus norvegicus</i> ]	66	46	207
734	3	673	512	gi1519085	phosphatidylcholine binding immunoglobulin heavy chain IGH variable region [ <i>Mus musculus</i> ]	66	60	162
740	1	3	317	gi11209272	argininosuccinate-lyase [ <i>Campylobacter jejuni</i> ]	66	47	315
764	1	310	747	gi1435296	alkaline phosphatase like protein [ <i>Lactococcus lactis</i> ] pir S39339 S39339 alkaline phosphatase-like protein - <i>Lactococcus lactis</i>	66	42	438
852	1	338	171	gi1536955	CG Site No. 361 [ <i>Escherichia coli</i> ]	66	43	168
886	1	3	158	gi1289272	ferrichrome-binding protein [ <i>Bacillus subtilis</i> ]	66	44	156
889	1	462	232	gi1813061	HCMVUL77 (AA 1-642) (Human cytomegalovirus)	66	66	231
893	1	2	247	gi1149008	putative [ <i>Helicobacter pylori</i> ]	66	45	246
900	1	1425	733	gi1580842	P3 [ <i>Bacillus subtilis</i> ]	66	51	693
906	2	2300	1473	gi1790945	aryl-alcohol dehydrogenase [ <i>Bacillus subtilis</i> ]	66	53	828
947	1	79	549	gi1410117	diaminopimelate decarboxylase [ <i>Bacillus subtilis</i> ]	66	47	471
950	1	1100	552	gi148713	orf145 [ <i>Staphylococcus aureus</i> ]	66	35	549
955	2	89	475	gi11204390	uridine kinase (uridine monophosphokinase) [ <i>Haemophilus influenzae</i> ]	66	50	387
981	2	1308	997	gi1457146	rhoptry protein [ <i>Plasmodium yoelii</i> ]	66	10	312
986	1	25	315	gi1305002	ORF_2356 [ <i>Escherichia coli</i> ]	66	31	291
1057	1	3	203	gi11303853	YggP [ <i>Bacillus subtilis</i> ]	66	40	201
1087	1	1	294	gi1575913	unknown [ <i>Saccharomyces cerevisiae</i> ]	66	53	294
1105	1	1	231	gi11045799	methylgalactoside permease ATP-binding protein [ <i>Mycoplasma genitalium</i> ]	66	46	231
1128	1	2	574	gi11001493	hypothetical protein [ <i>Synechocystis sp.</i> ]	66	46	573
1150	1	498	250	gi11499034	M. jannaschii predicted coding region MJ0255 [ <i>Methanococcus jannaschii</i> ]	66	40	249
1180	2	707	453	gi1215908	DNA polymerase (g43) [ <i>Bacteriophage T4</i> ]	66	46	255
1208	1	1123	587	gi1256653	DNA-binding protein [ <i>Bacillus subtilis</i> ]	66	58	537
1342	1	1	402	gi11208474	hypothetical protein [ <i>Synechocystis sp.</i> ]	66	53	402
1761	2	589	398	gi1215811	cell fiber protein [ <i>Bacteriophage T3</i> ]	66	50	192

TABLE 2



S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1981	1	499	251	gi 1045935	DNA helicase II (Mycoplasma genitalium)	66	40	249
2103	2	176	400	gi 929798	precursor for the major serotype surface antigens (Plasmodium aliciparum)	66	46	225
2141	1	373	188	gi 1256623	exodeoxyribonuclease (Bacillus subtilis)	66	38	186
2458	1	325	164	gi 1019410	unknown (Schizosaccharomyces pombe)	66	47	162
2505	1	468	235	gi 1510394	putative transcriptional regulator (Methanococcus jannaschii)	66	39	234
2525	1	558	280	gi 1000695	cytotoxin L (Clostridium sordellii)	66	44	279
2935	1	3	275	gi 765073	autolysin (Staphylococcus aureus)	66	47	273
3005	1	114	305	gi 1205784	heterocyst maturation protein (Haemophilus influenzae)	66	46	192
3048	1	80	277	gi 1303813	Yqew (Bacillus subtilis)	66	42	198
3071	1	1	189	gi 1070014	protein-dependent (Bacillus subtilis)	66	41	189
3081	1	404	225	gi 984212	unknown (Schizosaccharomyces pombe)	66	44	180
3090	2	580	386	gi 1204987	DNA polymerase III, alpha chain (Haemophilus influenzae)	66	48	195
3318	1	1	387	gi 1009366	Respiratory nitrate reductase (Bacillus subtilis)	66	49	387
3739	1	798	400	gi 1109684	Prox (Bacillus subtilis)	66	47	399
3796	1	402	202	gi 853760	acyl-CoA dehydrogenase (Bacillus subtilis)	66	60	201
1924	1	525	347	gi 563922	glucanase perimase (Bacillus licheniformis)	66	46	249
4240	1	3	350	gi 151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonii) pir[A44756][A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	66	51	348
4604	1	7	234	pir[A26713]BHC	hemocyanin subunit II - Atlantic horseshoe crab	66	46	228
4	9	8845	9750	gi 145646	cymr (Escherichia coli)	65	35	906
6	5	2708	3565	gi 887824	ORF_0310 (Escherichia coli)	65	47	858
13	1	1993	998	gi 143402	recombination protein (ttg start codon) (Bacillus subtilis) gi 1303923 RecN (Bacillus subtilis)	65	44	996
15	7	2493	3524	gi 1403126	csd gene product (Alcaligenes eutrophus)	65	38	1032
18	3	1908	1372	gi 349187	acyltransferase (Saccharomyces cerevisiae)	65	50	537
21	3	1467	2492	gi 149518	phosphoribosyl anthranilate transferase (Lactococcus lactis) pir[S35126][S35126 anthranilate phosphoribosyltransferase (EC 4.2.1.18) - Lactococcus lactis subsp. lactis	65	52	1026
25	4	3374	4312	gi 1502420	malonyl-CoA:Acyl carrier protein transacylase (Bacillus subtilis)	65	44	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
27	2	390	626	gi 1212729	YqjJ [Bacillus subtilis]	65	45	237
31	12	11040	10187	gi 509245	D-hydroxyacetone dehydrogenase [Lactobacillus delbrueckii]	65	41	654
38	24	19172	19528	gi 547	H-protein [Flavaria cronquistii]	65	41	357
44	2	790	1746	gi 405882	yekK [Escherichia coli]	65	46	957
44	12	9356	8832	gi 1205905	molybdenum cofactor biosynthesis protein [Haemophilus influenzae]	65	50	525
45	8	6635	7588	gi 491074	ApbA protein [Salmonella typhimurium]	65	46	954
51	2	580	1503	gi 580897	OppB gene product [Bacillus subtilis]	65	45	924
52	1	225	953	gi 1205518	NAD(P)H-flavin oxidoreductase [Haemophilus influenzae]	65	45	729
55	4	1339	1058	pir A4459 A444	troponin T beta TnT-5 - rabbit	65	41	282
67	9	7421	8272	gi 143607	sporulation protein [Bacillus subtilis]	65	42	852
73	5	4446	5375	gi 1204896	lysophospholipase L2 [Haemophilus influenzae]	65	37	930
74	1	954	478	gi 1204844	H. influenzae predicted coding region HI0594 [Haemophilus influenzae]	65	50	477
77	1	2	757	gi 1046082	M. genitalium predicted coding region MG372 [Mycoplasma genitalium]	65	46	756
77	2	795	1433	gi 1222116	perasease [Haemophilus influenzae]	65	37	639
81	3	4728	3454	gi 1001708	hypothetical protein [Synechocystis sp.]	65	49	1275
91	7	8548	8357	gi 1399263	cystathionine beta-lyase [Emerella nidulans]	65	40	192
98	3	1508	1988	gi 467423	unknown [Bacillus subtilis]	65	38	381
98	4	2250	2987	gi 467424	unknown [Bacillus subtilis]	65	45	738
102	3	2598	2119	gi 1511532	N-terminal acetyltransferase complex, subunit ARD1 [Methanococcus jannaschii]	65	39	480
102	4	3647	2862	gi 1204637	H. influenzae predicted coding region HI0388 [Haemophilus influenzae]	65	32	786
103	9	10851	9841	gi 142695	S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase Bacillus megaterium	65	47	1011
103	10	110439	10119	gi 710021	nitrite reductase (nirD) [Bacillus subtilis]	65	51	321
106	2	262	1140	gi 39881	ORP 311 (AA 1-311) [Bacillus subtilis]	65	44	879
109	5	3909	4268	gi 1204399	glucosamine-6-phosphate deaminase protein [Haemophilus influenzae]	65	44	360
109	10	7165	8595	gi 536955	ICD Site No. 361 [Escherichia coli]	65	41	1431

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
110	4	3688	3915	gi 407881	stringent response-like protein [Streptococcus equisimilis] pif[S9975]S9975 stringent response-like protein - Streptococcus equisimilis	65	45	228
110	5	3882	4295	gi 407880	orf1 [Streptococcus equisimilis]	65	50	414
110	6	4231	4380	gi 1139574	orf2 [Streptomyces griseus]	65	56	150
112	10	9218	8640	gi 1204571	H. influenzae predicted coding region H10318 [Haemophilus influenzae]	65	52	579
112	12	12049	11288	gi 710496	[transcriptional activator protein [Bacillus brevis]	65	32	762
125	1	2	202	gi 1151158	[repeat organellar protein [Plasmodium chabaudi]	65	39	201
126	1	3	422	gi 37589	[precursor [Homo sapiens]	65	46	420
127	11	10733	12658	gi 1064809	[homologous to sp. HTRA_ECOL1 [Bacillus subtilis]	65	41	1926
143	8	7543	7004	gi 216513	[mutator mutT (AT-OC transversion) [Escherichia coli]	65	56	540
145	5	3587	3838	gi 11209768	[D02_orf569 [Mycoplasma pneumoniae]	65	27	252
150	4	3482	2841	gi 1146225	[putative [Bacillus subtilis]	65	37	642
166	1	3858	1948	gi 148304	[beta-1,4-N-acetylglucosaminidase [Enterococcus hirae] pif[A42296]A42296 lysosyme 2 [EC 3.2.1.-] precursor - Enterococcus irae (ATCC 9790)	65	50	1911
188	6	3195	4178	gi 151943	[ORF3] putative [Rhodospirillum rubrum]	65	46	984
189	9	4982	4785	gi 158812	[ORF IV (AA 1-489) (Pigment mosaic virus)]	65	40	198
195	6	7900	5272	gi 145220	[alanyl-tRNA synthetase [Escherichia coli]	65	44	2637
195	7	10599	8104	gi 882711	[exonuclease V alpha-subunit [Escherichia coli]	65	38	2496
206	16	16896	18191	gi 408115	[ornithine acetyltransferase [Bacillus subtilis]	65	53	1296
217	4	3848	3215	gi 1205974	[5'-guanylate kinase [Haemophilus influenzae]	65	41	830
220	4	5265	3751	gi 580920	[ProD (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pif[S06048]S06048 probable form protein - bacillus subtilis sp[P13484]TACR_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LIPID-GLUCOSYLTRANSFERASE [EC 2.4.1.52] (TECHIOIC ACID BIOSYNTHESIS ROUTIN E)]	65	40	1515
236	5	2327	3709	gi 1146200	[DNA or RNA helicase, DNA-dependent ATPase [Bacillus subtilis]	65	46	1383
237	3	1902	2513	gi 149379	[HlsB [Lactococcus lactis]	65	46	612
241	4	4968	4195	gi 1205308	[ribonuclease H11 [EC 3.1.264] (RNASE H11) [Haemophilus influenzae]	65	50	774
252	1	1278	940	gi 1204589	[hypothetical protein [GB:U00022.9] [Haemophilus influenzae]	65	40	339
261	5	4780	3794	gi 145927	[fecD [Escherichia coli]	65	43	987

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
274	1	3	278	gi1496558	orfX (Bacillus subtilis)	65	42	276
301	2	982	815	gi1167418	unknown (Bacillus subtilis)	65	45	168
307	4	3586	2864	gi1070014	protein-dependent (Bacillus subtilis)	65	40	723
335	2	2286	1399	gi146913	N-acetylglucosamine transport protein (Escherichia coli) p1r1828895 WQEC2N phosphotransferase system enzyme II (EC .7.1.69), N-acetylglucosamine-specific - Escherichia coli sp P09323 PTAA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC I IABC OMPOENT (ELIA	65	50	888
338	5	4740	3170	gi11277029	biotin synthase (Bacillus subtilis)	65	49	951
343	3	1490	2800	gi1143264	membrane-associated protein (Bacillus subtilis)	65	48	1311
344	4	2761	2531	gi11050540	tRNA-glutamine synthetase (Lupinus luteus)	65	34	231
358	3	3421	3621	gi1146220	NAD+ dependent glycerol-3-phosphate dehydrogenase (Bacillus subtilis)	65	47	201
364	1	238	699	gi11340128	ORF1 (Staphylococcus aureus)	65	51	462
379	1	1	576	gi1143331	alkaline phosphatase regulatory protein (Bacillus subtilis) p1r147850 A27650 regulatory protein p10R - Bacillus subtilis sp P23545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN PHOR (EC 2.7.3.-)	65	40	576
379	3	3666	4346	gi1143268	ldihydrolypoamide transsuccinylase (odhB; EC 2.3.1.62) (Bacillus subtilis)	65	50	681
428	1	187	483	gi11420465	ORF YOR195W (Saccharomyces cerevisiae)	65	45	297
438	2	272	838	gi1143498	dkgS protein (Bacillus subtilis)	65	38	567
444	11	9280	10215	gi11204756	ribokinase (Haemophilus influenzae)	65	47	936
449	2	1241	1531	gi1599848	lha/M antiporter homolog (Lactococcus lactis)	65	41	291
478	2	1432	865	gi11045942	glycyl-tRNA synthetase (Hycoplasma genitalium)	65	39	588
479	1	1032	517	gi11498192	putative (Pseudomonas aeruginosa)	65	40	516
480	6	4312	5637	gi1415662	UDP-N-acetylglucosamine 1-carboxyvinyl transferase (Acinetobacter calcoaceticus)	65	48	1326
484	1	2	410	gi1146551	transmembrane protein (KdpD) (Escherichia coli)	65	44	429
499	1	54	932	gi1601456	reductase (Leishmania major)	65	53	879
505	1	914	459	gi11518853	lactA (Salmonella typhimurium)	65	39	456
571	2	1509	883	gi149399	open reading frame upstream glnE (Escherichia coli) tr1837754 S37754 (hypothetical) protein XE (glnE 5' region) - Escherichia coli	65	44	627
611	2	506	270	gi110961	RAP-2 (Plasmodium falciparum)	65	40	237

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
705	1	564	283	gi 710020	nitrite reductase (nirB) [Bacillus subtilis]	65	52	282
712	1	1	177	gi 289272	ferrichrome-binding protein [Bacillus subtilis]	65	37	177
712	2	196	354	gi 289272	ferrichrome-binding protein [Bacillus subtilis]	65	37	159
743	1	2	631	gi 310631	ATP binding protein [Streptococcus gordonii]	65	45	630
749	2	393	779	gi 167374	single strand DNA-binding protein [Bacillus subtilis]	65	29	387
762	1	1698	850	gi 160399	multidrug resistance protein [Plasmodium falciparum]	65	48	849
788	1	85	315	gi 1129096	unknown protein [Bacillus sp.]	65	35	231
850	1	1	408	gi 1006604	hypothetical protein [Synecococcus sp.]	65	37	408
90A	1	1	414	gi 1199546	2362 [Saccharomyces cerevisiae]	65	46	414
925	1	1	174	gi 1256653	DNA-binding protein [Bacillus subtilis]	65	54	174
1031	1	26	232	gi 238657	AppCycochrome d oxidase, subunit 1 homolog [Escherichia coli, K12, eptide, 514 aa]	65	47	207
1037	1	414	262	gi 1491813	gamma-glutamyltranspeptidase [Bacillus subtilis]	65	46	153
1051	1	348	175	gi 642655	unknown [Rhizobium meliloti]	65	34	174
1149	1	1399	752	gi 1162980	ribulose-5-phosphate 3-epimerase [Spinacia oleracea]	65	48	648
1214	1	881	495	gi 1205959	lactam utilization protein [Haemophilus influenzae]	65	45	387
1276	1	476	276	pir G55493 S354	site-specific DNA-methyltransferase SteI (EC 2.1.1.-) - Streptococcus sanguis	65	35	201
1276	2	900	577	gi 473794	ORF' [Escherichia coli]	65	34	124
2057	1	272	138	gi 633699	TrsH [Yersinia enterocolitica]	65	21	135
2521	1	336	169	gi 1045789	hypothetical protein (GB:U14003_76) [Mycoplasma genitalium]	65	41	168
2974	1	590	297	gi 152052	penicillinase-selective amidase [Rhodococcus sp.]	65	45	294
3031	1	306	154	pir J01024 J010	hypothetical 30K protein (bmrP140 5' region) - fruit fly (Drosophila melanogaster)	65	45	153
3069	1	3	278	gi 144906	product homologous to E.coli thiorodoxin reductase: J.Biol.Chem. 1988) 263:9015-9019, and to F52a protein of alkyl hydroperoxide oxidase from S. typhimurium: J.Biol.Chem. (1990) 265:10535-10540; pen reading frame A [Clostridium pasteurianum]	65	46	276
3146	1	282	142	gi 149315	ORF1 gene product [Bacillus subtilis]	65	47	141
3370	1	679	341	gi 1507711	indolepyruvate decarboxylase [Erwinia herbicola]	65	44	339

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
3546	1	1	303	gi 450688	hdm gene of EcoRII gene product [Escherichia coli] pir[S38437]S38437 hdm protein - Escherichia coli pir[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	65	42	303
3782	1	2	328	gi 166412	NADH-glutamate synthase (Medicago sativa)	65	42	327
3990	1	374	189	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]	65	53	186
4032	1	613	308	gi 1323127	ORF YG087C [Saccharomyces cerevisiae]	65	50	306
4278	2	726	364	gi 1197667	vitellogenin [Anolis pulchellus]	65	42	363
19	4	4259	5518	gi 145727	dead [Escherichia coli]	64	45	1280
19	6	7639	6926	gi 1016232	ycf27 gene product [Cyanophora paradoxa]	64	36	714
20	8	7053	6456	gi 765073	autolysin [Staphylococcus aureus]	64	47	600
31	13	12706	11537	gi 414009	lpe-85d gene product [Bacillus subtilis]	64	45	1170
33	4	2388	4364	gi 1204696	fructose-permease IIBC component [Haemophilus influenzae]	64	47	1977
36	3	1871	3013	gi 290503	glutamate permease [Escherichia coli]	64	40	1143
37	6	4065	4409	gi 39815	orf 2 gene product [Bacillus subtilis]	64	46	345
45	9	7852	8760	gi 1230585	nucleotide sugar epimerase [Vibrio cholerae O139]	64	53	909
51	3	1540	1899	gi 1303961	YQJ3 [Bacillus subtilis]	64	50	360
56	6	4793	3855	gi 457514	gltC [Bacillus subtilis]	64	45	939
56	24	30002	30247	gi 470331	similar to zinc fingers [Caenorhabditis elegans]	64	42	246
62	4	2759	2421	gi 642655	unknown [Rhizobium meliloti]	64	28	339
85	6	7178	6027	gi 457702	S-aminimidazole ribonucleotide-carboxylase [Pichia methanolica] pir[S39112]S39112 phosphoribosylaminoimidazole Carboxylase (EC .1.1.21) - yeast [Pichia methanolica]	64	46	1152
96	9	9251	10030	gi 1511513	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	64	42	780
100	1	1	600	gi 765073	autolysin [Staphylococcus aureus]	64	44	600
106	5	3668	4854	gi 466778	lysine specific permease [Escherichia coli]	64	46	987
123	2	838	554	gi 467484	unknown [Bacillus subtilis]	64	47	285
127	8	7514	7810	gi 210061	isotype-specific antigen [African horse sickness virus] pir[S27891]S27891 capsid protein VP2 - African horse sickness virus	64	28	297
131	7	7134	6721	gi 1511160	M. jannaschii predicted coding region MJ1163 [Methanococcus jannaschii]	64	46	414

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
142	5	5455	4817	gi 1173517	riboflavin synthase alpha subunit [Actinobacillus pleuropneumoniae]	64	44	639
143	1	709	356	pir A32950 A329	probable reductase protein - Leishmania major	64	52	354
149	10	3555	3295	gi 398151	major surface antigen MSG2 [Pneumocystis carinii]	64	44	261
154	4	3134	2307	gi 984587	DlnP [Escherichia coli]	64	50	828
161	5	3855	4880	gi 903304	ORF72 [Bacillus subtilis]	64	37	1026
165	1	33	791	gi 467483	unknown [Bacillus subtilis]	64	38	759
175	6	6355	4714	gi 1072398	phd gene product [Rhizobium meliloti]	64	42	1512
188	3	2042	2500	gi 001961	MHC class II analog [Staphylococcus aureus]	64	45	459
195	14	13667	13446	gi 396380	No definition line found [Escherichia coli]	64	47	222
206	15	16429	16938	gi 304134	argC [Bacillus stearothermophilus]	64	49	510
215	1	560	282	gi 143359	ORF 6 [Acetobacter vinelandii]	64	39	279
243	7	7818	6928	gi 414014	lpe-90d gene product [Bacillus subtilis]	64	49	891
258	2	1330	845	gi 664754	PI7 [Listeria monocytogenes]	64	38	486
259	1	462	232	gi 1499663	M. jannaschii predicted coding region M0817 [Methanococcus jannaschii]	64	52	231
263	6	6565	5567	gi 142828	aspartate semialdehyde dehydrogenase [Bacillus subtilis] sp Q04797 DHAS_DACSU ASPARTATE-SEMIALDEHYDE DEHYDROGENASE (EC 2.1.11) [ASA DEHYDROGENASE]	64	48	999
271	1	3	1163	gi 467091	hflX: B2235_C2_202 [Mycobacterium leprae]	64	44	1161
280	1	173	1450	gi 3303839	YqfR [Bacillus subtilis]	64	43	1278
293	1	2532	1267	gi 147345	primosomal protein n' [Escherichia coli]	64	45	1266
295	2	742	1488	gi 459266	Potential membrane spanning protein [Staphylococcus hominis] pir S42932 S42932 potential membrane spanning protein - taphylococcus hominis	64	39	747
301	5	1625	1446	gi 580835	lysine decarboxylase [Bacillus subtilis]	64	35	180
315	4	5064	3949	gi 141396	guinol oxidase [Bacillus subtilis]	64	45	1116
321	1	1264	635	gi 710496	transcriptional activator protein [Bacillus brevis]	64	41	630
333	5	4520	4239	gi 1114295	ORF2: putative 39 kDa protein [Listeria monocytogenes]	64	43	282
342	1	1	549	gi 142940	ftsA [Bacillus subtilis]	64	38	549
353	3	2878	2324	gi 537049	ORF_0470 [Escherichia coli]	64	44	555

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
379	2	827	3658	pir[S25295]A328	oxoglutarate dehydrogenase (lipoamide) (EC 1.2.4.2) - Bacillus subtilis	64	67	2832
404	6	4429	4839	pic[A36933]A369	diacylglycerol kinase homolog - Streptococcus mutans	64	35	411
407	1	2020	1133	gi 969026	OrfX [Bacillus subtilis]	64	41	888
425	1	1109	591	gi 1146177	phosphotransferase system glucose-specific enzyme II [Bacillus subtilis]	64	44	519
443	6	4082	4798	gi 147309	purine nucleoside-phosphorylase [Escherichia coli]	64	51	717
450	2	1035	1604	gi 106376	ORF_0162 [Escherichia coli]	64	38	570
470	5	1680	6107	gi 1369948	host interacting protein [Bacteriophage B1]	64	45	4428
486	4	1911	1471	gi 1205582	isermidine/putrescine transport system permease protein [Haemophilus influenzae]	64	35	441
497	1	2217	1159	sp P36929 FMU_E	FMU PROTEIN	64	38	1059
501	1	3	410	gi 142450	shc protein [Bacillus subtilis]	64	38	408
514	1	3	290	gi 1204496	H. influenzae predicted coding region H10238 [Haemophilus influenzae]	64	34	288
551	4	3162	3323	gi 1204511	bacterioferritin comigratory protein [Haemophilus influenzae]	64	41	162
603	4	759	956	gi 755823	NADH dehydrogenase P [Streptococcus americana]	64	35	198
653	2	940	746	gi 121234	dicarboxylic amino acids DapB permease [Saccharomyces cerevisiae]	64	41	195
660	3	3801	2257	sp P46133 YDIAU_	HYPOTHETICAL PROTEIN IN OCT 5' REGION [FRAGMENT]	64	39	1545
695	1	11	502	gi 1001383	hypothetical protein [Synecocystis sp.]	64	41	492
702	1	3	752	gi 142865	DNA primase [Bacillus subtilis]	64	46	750
826	1	1	339	gi 971336	arginyl tRNA synthetase [Bacillus subtilis]	64	50	339
838	1	1831	917	gi 1354775	pfosR [Treponema pallidum]	64	41	915
864	3	675	944	gi 39833	cyclomaltodextrin glucanotransferase [Bacillus stearothermophilus] i39835	64	47	270
887	1	3	677	gi 153002	enterotoxin type E precursor [Staphylococcus aureus] pir A28179 A28179 enterotoxin E precursor - Staphylococcus aureus sp P1993 ETXE_STAAU ENTEROTOXIN TYPE E PRECURSOR (SEE)	64	46	675
928	2	1172	963	gi 111976	fibrinogen-binding protein [Staphylococcus aureus] pir S34270 S34270 fibrinogen-binding protein - Staphylococcus aureus	64	41	210
1049	2	800	606	gi 1049115	Pap60 [Bacillus subtilis]	64	42	195
1067	2	999	748	gi 1151072	Hda precursor [Haemophilus ducreyi]	64	50	252

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1120	1	50	202	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	64	30	153
1125	1	751	377	gi 581648	epiB gene product [Staphylococcus epidermidis]	64	44	375
1688	1	402	214	gi A01365 TVMS	transforming protein K-tes - mouse	64	47	189
2472	1	2	358	gi 487282	Ne <sup>+</sup> -ATPase subunit J [Enterococcus hirae]	64	36	357
2889	1	520	356	gi 304134	argC [Bacillus stearothermophilus]	64	50	165
3013	1	630	352	gi 551699	cytochrome oxidase subunit I [Bacillus firmus]	64	51	279
3014	1	546	274	gi 1204349	hypothetical protein (GB:G8:090212_3) [Haemophilus influenzae]	64	50	273
3197	1	613	308	gi 1009366	respiratory nitrate reductase [Bacillus subtilis]	64	46	306
3303	1	90	362	gi 1107839	alginate lyase [Pseudomonas aeruginosa]	64	43	273
3852	2	82	288	gi 216746	D-lactate dehydrogenase [Lactobacillus plantarum]	64	42	207
3868	1	1	312	gi 149435	putative [Lactococcus lactis]	64	48	312
3918	1	660	331	gi 5532	acetyl-CoA acyltransferase [Yarrowia lipolytica]	64	46	330
4000	1	112	378	gi 944688	unknown [Saccharomyces cerevisiae]	64	44	287
4009	1	81	368	gi 39372	grsB gene product [Bacillus brevis]	64	41	288
4166	1	2	349	gi 149435	putative [Lactococcus lactis]	64	46	348
4366	1	2	307	gi 216267	ORF2 [Bacillus megaterium]	64	44	306
4457	1	2	400	gi 1197667	vitallogenin [Anolis pulchellus]	64	43	399
11	3	1539	2438	gi 438228	ORF C [Staphylococcus aureus]	63	32	900
24	7	5611	5423	gi 1369943	a1 gene product [Bacteriophage 81]	63	34	189
29	1	1	390	gi 467441	expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis] gi 467441	63	43	390
					expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis]			
					expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis]			
31	6	6329	5712	gi 496943	ORF [Saccharomyces cerevisiae]	63	47	618
44	123	14669	15019	gi A04446 QOEC	hypothetical protein P-92 - Escherichia coli	63	36	351
48	6	4403	6250	gi 43498	pyruvate synthase [Halobacterium halobium]	63	42	1848
50	5	3869	4738	gi 413967	ipa-43d gene product [Bacillus subtilis]	63	43	870
53	6	6764	5742	gi 474176	regulator protein [Staphylococcus xylosus]	63	49	1023

TABLE 2

3. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match	Gene name	% sim	% ident	length (nt)
56	14	15880	17607	gt 467409	DNA polymerase III subunit [Bacillus subtilis]	63	44	1728
57	11	7945	7376	gt 137036	ORF_0158 [Escherichia coli]	63	39	570
62	3	2479	2114	gt 142656	unknown [Rhizobium meliloti]	63	41	366
70	8	6562	7353	gt 139821	PhoC [Rhizobium meliloti]	63	46	792
75	2	223	927	gt 149376	HsdC [Lactococcus lactis]	63	45	705
78	5	4912	4403	gt 413950	lpa-26d gene product [Bacillus subtilis]	63	42	510
91	5	9076	7220	gt 466997	lactH2; B2126-C1_157 [Mycobacterium leprae]	63	41	1857
91	8	10566	9448	gt 1204344	cytostathionine gamma-synthase [Haemophilus influenzae]	63	45	1119
120	1	21	1508	gt 882657	sulfite reductase (NADPH) flavoprotein beta subunit [Escherichia coli]	63	46	1488
120	4	2722	4125	gt 665994	hypothetical protein [Bacillus subtilis]	63	34	1404
127	7	6064	7366	gt 140182	murE gene product [Bacillus subtilis]	63	44	1303
149	6	2321	2106	gt 148503	dnaK [Eryipelothrix rhusiopathiae]	63	40	216
149	26	10445	10170	gt 4870	ORF 2, has similarity to DNA polymerase [Saccharomyces kluyveri] r S15961 S15961 hypothetical protein 2 - yeast [Saccharomyces yverli] plasmid pSKL	63	42	276
164	2	507	1298	gt 145476	CDP-diglyceride synthetase [Escherichia coli]	63	44	792
166	6	9009	8164	gt 151012	fructooligosaccharide 11 [Rhodospirillum rubrum]	63	41	1746
169	4	1704	1886	gt 152886	elongation factor Ts (tsf) [Spiroplasma citri]	63	48	183
188	5	3145	2951	gt 1334547	GlyC1 114 grp 18 protein [Podospira anserina]	63	42	195
195	13	11767	12804	gt 606100	ORF_0335 [Escherichia coli]	63	40	1018
201	2	607	2283	gt 433534	arginyl-tRNA synthetase [Corynebacterium glutamicum] pir A49936 A49936 arginine-tRNA ligase (EC 6.1.1.19) - oryzae [Corynebacterium glutamicum]	63	46	1677
206	14	15893	16489	gt 580828	N-acetyl-glucamate-gamma-semialdehyde dehydrogenase [Bacillus subtilis]	63	49	597
220	5	7769	5766	gt 216334	lscA protein [Bacillus subtilis]	63	42	2004
221	1	74	907	gt 677945	AppA [Bacillus subtilis]	63	42	834
227	3	944	1708	gt 1510558	isobutyrate synthase [Methanococcus jannaschii]	63	46	765
261	2	804	1070	gt 486511	ORF_YK054c [Saccharomyces cerevisiae]	63	45	267
269	2	3606	1960	gt 148221	DNA-dependent ATPase, DNA helicase [Escherichia coli] pir J50137 J50137 recQ protein - Escherichia coli	63	42	1647

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
278	8	7417	8176	gi 699273	cystathionine gamma-synthase [Mycobacterium leprae] sp P46807 MEYS_MYCLE CYSTATHIONINE GAMMA-SYNTASE [EC 4.2.99.9] O-SUCCINYLMOSERINE (THIOL)- LYASE)	63	41	1242
287	2	738	1733	gi 405133	[putative [Bacillus subtilis]	63	38	996
295	1	2	748	gi 123983	hypothetical protein [Bacillus subtilis]	63	41	747
328	3	2148	3134	gi 45302	carrier protein (AA 1 - 437) [Paenomonas aeruginosa] ir S11497 S11497 branched-chain amino acid tr: :port protein brab - eudomonas aeruginosa	63	36	987
362	2	1226	1216	sp P51336 SERA_	D-3-PHOSPHOGLYCERATE DEHYDROGENASE [EC 1.1.1.95] (PQDH) .	63	38	411
404	1	326	1051	gi 1303816	YqeZ [Bacillus subtilis]	63	35	726
405	3	2101	1715	gi 1303914	YqhY [Bacillus subtilis]	63	42	387
406	3	451	227	gi 142152	sulfate permease (g9 start codon) [Synecococcus PCC6301] pir A3030 GRYCS7 sulfate transport protein - Synecococcus sp. PCC 7942)	63	43	225
415	2	1048	2718	gi 1205402	transport ATP-binding protein [Haemophilus influenzae]	63	41	1671
426	4	3575	2679	gi 193268	29-kilodalton protein [Streptococcus pneumoniae] sp P42362 P29K_STPM 29 KD MEMBRANE PROTEIN IN PSAA 5'REGION ORF1).	63	39	897
505	3	1347	2195	gi 1618999	orf4 [Lactobacillus sake]	63	40	849
507	1	2	574	gi 146917	comK [Bacillus subtilis, E26, Peptide, 192 aa]	63	35	573
562	2	146	1084	gi 143985	nifs-like gene [Lactobacillus delbrueckii]	63	45	939
675	1	427	215	gi 1510994	isoleucine aminotransferase [Methanococcus jannaschii]	63	29	213
686	1	3	230	gi 1517356	nitrate reductase (NADH) [Lotus japonicus]	63	52	228
701	1	3	392	gi 881940	NorQ protein [Paracoccus denitrificans]	63	41	390
720	1	2	400	gi 47168	open reading frame [Streptomyces lividans]	63	35	399
779	1	571	287	gi 1261932	unknown [Mycobacterium tuberculosis]	63	41	285
907	1	22	321	gi 149445	ORF1 [Lactococcus lactis]	63	27	300
972	1	794	399	gi 1511235	M. jannaschii predicted coding region M2232 [Methanococcus jannaschii]	63	27	396
1085	1	1154	618	gi 1204277	hypothetical protein (CB:U00019_14) [Haemophilus influenzae]	63	38	537
1094	1	3	542	gi 790943	urea amidolyase [Bacillus subtilis]	63	39	540
1108	1	3	482	pir S4982 S498	regulation protein - Bacillus subtilis	63	44	480
1113	1	1231	617	gi 493017	endocarditis specific antigen [Enterococcus faecalis]	63	45	615

TABLE 2

TABLE 2

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	naidb gene name	% sim	% ident	length (nt)
1300	1	3	695	sp P33940 YQHJ_	HYPOTHETICAL 54.3 NO PROTEIN IN ECO-ALKB INTERGENIC REGION.	63	46	693
1325	1	1	204	gi 928989	pi100 protein [Borrelia burgdorferi]	63	30	204
1814	1	3	245	gi 1303914	YqHY [Bacillus subtilis]	63	34	243
2021	1	498	250	pir C33496 C334	hisc homolog - Bacillus subtilis	63	46	249
2325	1	2	193	gi 436132	product is similar to TnpA of transposon Tn554 from Staphylococcus aureus [Clostridium butyricum]	63	40	192
2335	1	1	195	gi 1184298	flagellar MS-ring protein [Borrelia burgdorferi]	63	47	195
2406	1	451	227	gi 1041785	phoptry protein [Plasmodium yoelii]	63	33	225
2561	2	136	360	gi 312443	carbamoyl-phosphate synthase (glutamine-hydrolysing) [Bacillus aldolyticus]	63	52	225
2565	1	1	402	gi 1407784	orf-1; novel antigen [Staphylococcus aureus]	63	50	402
2587	1	583	292	gi 1224049	amidase [Moraxella catarrhalis]	63	35	291
2594	1	266	135	gi 836646	phosphoribosylformimino-praic ketoisomerase [Rhodobacter phaeoideae]	63	51	132
3043	1	440	252	gi 1480237	phenylacetaldehyde dehydrogenase [Escherichia coli]	63	40	189
3078	1	609	400	gi 1487982	intrinsic membrane protein [Mycoplasma hominis]	63	36	210
3139	1	2	217	gi 439126	glutamate synthase (NADPH) [Azospirillum brasilense] pir A49916 A49916 glutamate synthase (NADPH) [EC 1.4.1.13] - Azospirillum brasilense	63	47	216
3625	1	793	198	gi 623073	ORF160; putative [Haemophilus influenzae (H1N2)]	63	48	196
3658	1	1	399	gi 1303697	YrKA [Bacillus subtilis]	63	37	399
3659	1	3	395	gi 1256135	YbbF [Bacillus subtilis]	63	48	393
3783	1	720	361	gi 1256902	Pyruvate decarboxylase isozyme 2 (Swiss Prot. accession number P16467) [Saccharomyces cerevisiae]	63	34	360
3900	1	338	171	sp P0537 AMYB_	BETA-AMYLAASE [EC 3.2.1.2] (1,4-ALPHA-D-GLUCAN MALTOHYDROLASE).	63	54	168
4309	1	3	176	pir A37967 A379	neuronal cell adhesion molecule Ng-CAM precursor - chicken	63	57	174
4367	1	1	195	gi 131932	Par6p gene product [Pichia pastoris]	63	30	195
4432	1	1	312	gi 151259	hmg-CoA reductase [EC 1.1.1.88] (Pseudomonas nevalonii) pir A44756 A44756 hydroxymethylglucaryl-CoA reductase [EC 1.1.1.88] Pseudomonas sp.	63	51	312
4468	1	6	308	gi 296464	ATPase [Lactococcus lactis]	63	36	303
33	3	1411	2400	gi 1513675	tagatase 6-P kinase [Streptococcus mutans]	62	44	990
36	9	5985	6218	gi 1490521	hUHS3 [Homo sapiens]	62	51	234

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
37	1	2	721	gi1107531	ceuE gene product (Campylobacter coli)	62	33	720
38	15	10912	11589	gi1272058	H. influenzae predicted coding region H1M1279 (Haemophilus influenzae)	62	38	678
38	25	19526	20329	gi1695280	ORF2 (Alcaligenes eutrophus)	62	41	804
57	2	2523	1780	gi1171234	orf1 (Haemophilus influenzae)	62	55	744
57	9	6646	6350	gi1508174	EF1B domain of PTS-dependent Gal transport and phosphorylation Escherichia coli	62	35	297
58	1	2	559	gi1755152	highly hydrophobic integral membrane protein (Bacillus subtilis) sp142953 TAGC_BACSU TEICHOIC ACID TRANSLOCATION PERMEASE PROTEIN AGC.	62	34	558
67	10	8250	9014	gi1470683	Shows similarity with ATP-binding proteins from other ABC-transport porins, Swiss Prot Accession Numbers P24137, P08007, P04285, P24136 Escherichia coli	62	31	765
69	8	8315	7494	gi146816	actA 4 gene product (Streptomyces coelicolor)	62	44	822
80	3	1793	1320	gi139593	UDP-N-acetylglucosamine-6-phosphate ligase (Bacillus subtilis)	62	43	474
87	7	7034	9205	gi1217191	5'-nucleotidase precursor (Vibrio parahaemolyticus)	62	38	2172
100	3	4051	3089	gi1511047	phosphoglycerate dehydrogenase (Methanococcus jannaschii)	62	42	963
102	1	2	520	gi153655	mismatch repair protein (Streptococcus pneumoniae) pir[C28667] C28667 DNA mismatch repair protein hexA - Streptococcus pneumoniae	62	34	519
112	2	466	1068	gi153741	ATP-binding protein (Streptococcus mutans)	62	37	603
114	7	6855	7562	gi1204866	L-fucose operon activator (Haemophilus influenzae)	62	38	708
116	4	6823	5633	gi1677947	AppC (Bacillus subtilis)	62	37	1191
124	8	6855	6004	gi1853777	product similar to E.coli PRA2 protein (Bacillus subtilis) pir 553438 553438 ynfK protein - Bacillus subtilis sp145873 HNMK_BACSU POSSIBLE PROTOPHYRINOGEN OXIDASE (EC 3.3.-)	62	44	852
148	1	24	554	gi1467456	unknown (Bacillus subtilis)	62	30	531
149	20	7591	6725	gi11205807	replicative DNA helicase (Haemophilus influenzae)	62	41	867
163	3	1303	1153	gi140067	X gene product (Bacillus sphaericus)	62	42	351
164	15	14673	15632	gi142219	P15 gene product (AA 1 - 314) (Escherichia coli)	62	38	960
165	2	1166	1447	gi1403936	phenylalanyl-tRNA synthetase alpha subunit (Gly294 variant) unidentified cloning vector	62	38	282
166	2	2084	5089	gi1308861	GTC start codon (Lactococcus lactis)	62	44	3006
171	1	1225	614	gi1046053	hypothetical protein (SP-P12049) (Mycoplasma genitalium)	62	41	612

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
183	1	2521	1310	gi 143045	hemY [Bacillus subtilis]	62	45	1212
200	1	3	956	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	62	32	954
237	2	935	1966	gi 141695	hisC protein [Escherichia coli]	62	44	1032
261	3	4008	2605	gi 143121	ORF A1 putative [Bacillus firmus]	62	42	1404
299	8	4477	4719	gi 1467441	expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed [Bacillus subtilis] gi 1467441 expressed at the end of exponential growth under conditions in which the enzymes of the TCA cycle are repressed Bacil	62	47	243
304	6	5018	3819	gi 153015	PemA protein [Staphylococcus aureus]	62	43	1200
324	1	2	262	gi 142717	cytochrome aa3 controlling protein [Bacillus subtilis] pir[AJ3960]AJ3960 cta protein - Bacillus subtilis sp P12946 CTAA_BACSU CYTOCHROME AA3 CONTROLLING PROTEIN	62	30	261
325	2	269	1207	gi 1581088	methionyl-tRNA formyltransferase [Escherichia coli]	62	39	939
332	6	4894	4631	gi 1499960	uridine 5'-monophosphate synthase [Methanococcus jannaschii]	62	36	264
355	1	2	370	gi 145925	fecB [Escherichia coli]	62	32	369
365	8	6628	6804	gi 133943	lipA-19d gene product [Bacillus subtilis]	62	54	177
369	2	2744	1626	pir[AJ3577]AJ35	regulatory protein pfor - Clostridium perfringens	62	42	1119
370	1	34	264	gi 140665	beta-glucosidase [Clostridium thermocellum]	62	37	231
415	3	2709	1176	gi 1205401	transport ATP-binding protein [Haemophilus influenzae]	62	35	468
429	1	1578	790	gi 1046024	Na+ ATPase subunit J [Mycoplasma genitalium]	62	40	789
444	2	704	1369	gi 1581510	modulation gene: integral membrane protein; homology to Rhizobium eguminosarum nodI [Rhizobium loti]	62	37	666
477	2	751	1869	pir[AJ8440]AJ84	ring-infected erythrocyte surface antigen 2, RESA-2 - Plasmodium falciparum	62	44	1119
485	1	241	1707	gi 117934	betaine aldehyde dehydrogenase [Beta vulgaris]	62	43	1467
487	3	1141	1311	gi 149445	ORF1 [Lactococcus lactis]	62	31	171
494	2	1134	1313	gi 1166835	ribulose biphosphate carboxylase/oxygenase activase [Arabidopsis thaliana]	62	37	180
518	1	193	882	gi 153491	O-methyltransferase [Streptomyces glaucus]	62	39	690
534	2	369	2522	gi 1480429	putative transcriptional regulator [Bacillus stearothermophilus]	62	35	2154
551	6	4371	4820	gi 111113	ferric uptake regulation protein [Campylobacter jejuni]	62	37	450
574	1	1	570	gi 153000	enterotoxin B [Staphylococcus aureus]	62	43	570

TABLE 2

S aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
590	2	344	1171	gi 40367	ORF1 [Clostridium acetobutylicum]	62	37	828
655	1	396	830	gi 147195	phb protein [Escherichia coli]	62	44	435
656	1	2	478	gi 1205451	cell division inhibitor [Haemophilus influenzae]	62	36	477
676	1	692	348	gi 1511613	methyl coenzyme M reductase system, component A2 [Methanococcus jannaschii]	62	36	345
687	1	493	248	gi 49272	Asparaginase [Bacillus licheniformis]	62	48	246
700	2	267	944	gi 1205822	hypothetical protein (GB:X75627.4) [Haemophilus influenzae]	62	40	678
840	2	1715	1041	gi 1045865	M. genitalium predicted coding region NG181 [Mycoplasma genitalium]	62	36	675
864	4	898	1491	gi 1144332	deoxyuridine nucleotidohydrolase [Homo sapiens]	62	38	594
916	1	35	400	gi 413931	ipa-7d gene product [Bacillus subtilis]	62	45	366
1071	1	1	771	gi 1510649	aspartokinase I [Methanococcus jannaschii]	62	40	771
1084	1	19	609	gi 688011	AQX-1 antigen [human, infertile patient, testis, peptide, 505 aa]	62	39	591
1103	1	3	203	gi 581261	ORF homologous to E. coli, mob [Mycobacterium aurantiacum] p1r[S14030]S14030	62	51	201
					Hypothetical protein - Mycobacterium aurantiacum (fragment)			
1217	1	463	233	gi 460025	ORF2, putative [Streptococcus pneumoniae]	62	41	231
1533	1	644	414	gi 413968	ipa-4d gene product [Bacillus subtilis]	62	48	231
1537	1	3	257	gi 1510641	alanyl-tRNA synthetase [Methanococcus jannaschii]	62	29	255
2247	1	3	161	gi 485956	myc gene product [Proteus mirabilis]	62	45	159
2386	1	3	245	gi 285708	nontoxic component [Clostridium botulinum]	62	31	243
2484	1	331	167	gi 142092	DNA-repair protein (recA) [Anabaena variabilis]	62	35	165
2490	1	798	400	gi 581648	epib gene product [Staphylococcus epidermidis]	62	42	399
3016	1	596	300	gi 710022	uroporphyrinogen III [Bacillus subtilis]	62	51	297
3116	1	1	213	gi 466883	nifs; B1496_C2_19 [Mycobacterium leprae]	62	44	213
3297	1	823	413	gi 475715	acetyl coenzyme A acetyltransferase (thiolase) [Clostridium acetobutylicum]	62	42	411
3609	1	31	276	gi 1408501	homologous to N-acetyl-L-amino acid amidohydrolase of Bacillus stearothermophilus [Bacillus subtilis]	62	48	246
3665	2	584	402	gi 151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mavaloni] p1r[A44756]A44756	62	40	183
					hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.			
3733	1	3	374	gi 1351197	thioredoxin reductase [Cubacterium acidaminophilum]	62	42	372

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3898	1	1	237	gi 153675	tagatase 6-P kinase [Streptococcus mutans]	62	45	237
4027	1	283	143	gi 310705	homologue to gene 30 (aa 1-59); putative [Sovine herpesvirus 4]	62	43	141
4109	1	727	365	gi 41748	hdm protein (AA 1-520) [Escherichia coli]	62	45	363
4303	1	1	303	gi 1303813	Yqew [Bacillus subtilis]	62	43	303
4380	1	530	267	gi 1235684	mevalonate pyrophosphate decarboxylase [Saccharomyces cerevisiae]	62	55	264
4494	1	2	256	gi 510692	enterotoxin H [Staphylococcus aureus]	62	34	255
4598	1	411	223	gi 763513	ORF4; putative [Streptomyces violaceoruber]	62	45	189
4624	1	1	222	gi 41748	hdm protein (AA 1-520) [Escherichia coli]	62	45	222
5	5	4288	3912	gi 928831	ORF95; putative [Lactococcus lactis phage BK5-T]	61	36	357
11	1	320	162	pir C33356 C333	prothymosin alpha homolog (clone 32) - human (fragment)	61	33	159
16	11	10991	11918	gi 1205391	hypothetical protein (SP:033995) [Haemophilus influenzae]	61	44	948
32	1	281	801	gi 1066504	exo-beta 1.1 glucanase [Cochliobolus carbonum]	61	50	519
38	3	616	1107	gi 1510864	glutamine transport ATP-binding protein Q [Methanococcus jannaschii]	61	41	492
45	4	3082	4018	gi 1109686	ProX [Bacillus subtilis]	61	45	957
48	8	7118	7504	gi 498839	ORF2 [Clostridium perfringens]	61	33	387
51	9	4605	5570	gi 388269	trac [Plasmodium falciparum]	61	42	966
60	6	1689	2243	gi 1205893	hypothetical protein (DB:000011_3) [Haemophilus influenzae]	61	32	555
62	9	5559	5122	gi 854656	Na/H antiporter system ORF2 [Bacillus alcalophilus]	61	38	438
67	5	4330	5846	gi 466612	lnika [Escherichia coli]	61	36	1317
74	2	2400	1504	gi 1204846	carbamate kinase [Haemophilus influenzae]	61	40	897
85	1	2198	1101	gi 1498756	amidophosphoribosyltransferase PurF [Rhizobium etli]	61	41	1098
86	4	1995	1582	gi 149931	M. jannaschii predicted coding region MJ1083 [Methanococcus jannaschii]	61	44	414
97	1	74	649	gi 1518679	orf [Bacillus subtilis]	61	44	576
99	2	2454	1990	gi 413958	lipa-34d gene product [Bacillus subtilis]	61	18	465
124	7	6223	5123	gi 556881	[Similar to Saccharomyces cerevisiae SUA5 protein [Bacillus subtilis] pir S49358 S49358 ipc-29d protein - Bacillus subtilis sp P3153 YMLC_BACSU HYPOTHETICAL 37.0 KD PROTEIN IN SPOTIR-GLYC INTERGENIC REGION.]	61	46	1101
125	4	1668	2531	gi 1491643	ORFA gene product [Chloroflexus aurantiacus]	61	43	864

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	blast gene name	% sim	% ident	length (nt)
132	1	1250	627	pir P00259 P002	hypothetical protein I - Enterococcus faecalis plasmid pAM-beta-1 (fragment)	61	43	624
149	9	1617	1075	gi 1144332	deoxyuridine nucleotidohydrolase [Homo sapiens]	61	40	543
149	22	8690	7869	gi 1160047	p101/acidic basic repeat antigen (Plasmodium falciparum) p1r A29232 A29232 101K malaria antigen precursor - Plasmodium falciparum (strain Camp)	61	35	822
166	3	1915	2361	gi 1499694	HIT protein, member of the HIT-family [Methanococcus jannaschii]	61	41	447
171	9	9675	7948	gi 467436	similar to SpvB [Bacillus subtilis]	61	38	1728
174	3	1042	2340	gi 216374	glucaryl 7-ACA acylase precursor [Bacillus laterosporus]	61	49	1299
190	4	5034	4111	gi 409286	bmrU [Bacillus subtilis]	61	37	924
216	1	2	190	gi 415861	eukaryotic initiation factor 2 beta (eIF-2 beta) [Oryctolagus uniculus]	61	29	189
227	7	4161	5048	gi 216341	ORF for methionine amino peptidase [Bacillus subtilis]	61	41	888
238	4	1959	3047	gi 809543	ChRC protein [Erwinia chrysanthemi]	61	38	1089
247	1	2	694	gi 537231	ORF_579 [Escherichia coli]	61	38	693
247	2	678	1034	gi 142226	chvD protein [Agrobacterium tumefaciens]	61	40	357
257	2	3523	2627	gi 699379	glvr-1 protein [Mycobacterium leprae]	61	40	897
268	2	3419	3051	gi 40364	ORF1 [Clostridium acetobutylicum]	61	41	369
275	4	4621	4827	gi 1204848	hypothetical protein (GP:H87049.57) [Haemophilus influenzae]	61	36	207
277	1	1	1845	gi 794897	beta-N-acetylhexosaminidase [Streptococcus pneumoniae] p1r A56390 A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - Streptococcus pneumoniae	61	45	1845
278	9	8003	7032	gi 467462	cysteine synthetase A [Bacillus subtilis]	61	43	972
278	10	9878	8535	gi 1205919	Na+ and Cl- dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]	61	38	1344
283	1	1	366	gi 755607	polyA polymerase [Bacillus subtilis]	61	36	366
288	2	1918	1496	gi 388108	cell wall enzyme [Enterococcus faecalis]	61	43	423
291	1	86	334	gi 454265	FBP3 [Petunia hybrida]	61	38	249
318	1	1104	694	gi 290531	similar to beta-glucoside transport protein [Escherichia coli] sp P11451 PT18-ECOLI PTS SYSTEM, ARABUTIN-LIKE II B COMPONENT PHOSPHOTRANSFERASE ENZYME II, B COMPONENT (EC 2.7.1.69)	61	47	411
330	2	1912	1190	gi 1001805	hypothetical protein [Synechocystis sp.]	61	41	723

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
385	2	1513	1025	gi 533098	[dnaD protein (Bacillus subtilis)]	61	42	489
426	1	794	399	gi 1303853	[yqef (Bacillus subtilis)]	61	44	396
438	3	810	1421	gi 1393660	[abaA2 (Streptomyces coelicolor)]	61	36	612
454	1	1580	792	gi 733522	[phosphatidylinositol-4,5-diphosphate 3-kinase (Dictyostelium discoideum)]	61	30	789
464	2	784	560	gi 1123120	[CS387.5 gene product (Caenorhabditis elegans)]	61	38	225
470	8	6077	7357	gi 623073	[ORF360; putative (Bacteriophage LL-H)]	61	47	1281
509	1	554	279	gi 467484	[unknown (Bacillus subtilis)]	61	45	276
555	3	1916	1296	gi 141800	[anthranilate synthase glutamylase (Acinetobacter baumannii)]	61	42	621
569	1	1711	857	gi 467090	[82235_C2_195 (Mycobacterium leprae)]	61	47	855
585	2	961	803	sp P36686 SURE_	[SURVIVAL PROTEIN SURE HOMOLOG (FRAGMENT)]	61	33	159
592	3	1694	1422	gi 1221602	[immunity repressor protein (Haemophilus influenzae)]	61	32	273
603	1	43	357	gi 507738	[ilap (Vibrio parahaemolyticus)]	61	33	315
669	1	2467	1235	gi 1146243	[22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative (Bacillus subtilis)]	61	37	1233
675	3	805	1101	gi 403373	[glycerophosphoryl diester phosphodiesterase (Bacillus subtilis)]	61	36	297
703	1	1656	829	gi 537181	[ORF_470 (Escherichia coli)]	61	32	828
728	1	1628	816	gi 606281	[DNA polymerase I (Bacillus stearothermophilus)]	61	39	813
823	1	61	318	gi 709992	[hypothetical protein (Bacillus subtilis)]	61	38	258
856	2	2313	1567	gi 609310	[portal protein gp3 (Bacteriophage HK97)]	61	40	747
923	1	1081	542	gi 143213	[putative (Bacillus subtilis)]	61	38	540
1124	1	59	370	gi 1107541	[C33D9.8 (Caenorhabditis elegans)]	61	26	312
1492	1	348	276	gi 406397	[unknown (Mycoplasma genitalium)]	61	32	273
1602	1	46	318	gi 733522	[phosphatidylinositol-4,5-diphosphate 3-kinase (Dictyostelium discoideum)]	61	34	273
2500	1	577	290	gi 1045964	[hypothetical protein (CBU14003_297) (Mycoplasma genitalium)]	61	31	288
2968	1	2	808	gi 397526	[clumping factor (Staphylococcus aureus)]	61	55	807
3076	1	3	248	gi 149373	[ORF 1 (Lactococcus lactis)]	61	41	246

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
3609	2	207	401	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus stearothermophilus</i> [Bacillus subtilis]	61	39	195
3662	1	1477	740	gi 1303813	Yqew [Bacillus subtilis]	61	42	738
3672	1	2	442	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir A56390 A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	50	441
3724	1	2	220	gi 1009366	Respiratory nitrate reductase [Bacillus subtilis]	61	41	219
3728	1	3	398	gi 677943	AppD [Bacillus subtilis]	61	46	396
3884	1	3	401	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir A56390 A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	47	399
3971	1	3	383	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir A56390 A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	45	381
4038	1	661	359	gi 1339970	large subunit of NAM-dependent glutamate synthase (Plectonasma boryanum)	61	24	303
4041	1	546	274	gi 413953	lga-29d gene product [Bacillus subtilis]	61	48	273
4047	1	1	402	gi 528991	unknown [Bacillus subtilis]	61	42	402
4102	1	1	345	gi 976025	HlsA [Escherichia coli]	61	46	345
4155	1	1	336	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir A56390 A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	50	336
4268	1	463	233	gi 450688	hsdM gene of EcoPrt1 gene product [Escherichia coli] pir S38437 S38437 hsdH protein - Escherichia coli pir S09629 S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	61	38	231
4374	1	542	273	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir A56390 A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	50	270
4389	1	2	172	gi 147516	ribokinase [Escherichia coli]	61	35	171
4621	1	2	268	gi 784897	beta-N-acetylhexosaminidase (Streptococcus pneumoniae) pir A56390 A56390 mannosyl-glycoprotein ndo-beta-N-acetylglucosaminidase (EC 3.2.1.96) precursor - streptococcus pneumoniae	61	47	267
4663	1	27	227	gi 976025	HlsA [Escherichia coli]	61	50	201
4	6	663	5536	gi 1408501	homologous to N-acyl-L-amino acid amidohydrolase of <i>Bacillus stearothermophilus</i> [Bacillus subtilis]	60	43	1128

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
11	6	3426	3725	gi 410748	ring-infected erythrocyte surface antigen (Plasmodium falciparum) p125526125526 ring-infected erythrocyte surface antigen precursor - Plasmodium falciparum (strain FC27/Papua New Guinea) sp P13830 RESA_PLAFF	60	24	300
11	14	11035	10313	gi 1217651	carboxyl reductase (NADPH) (Rattus norvegicus)	60	28	723
16	12	11917	12930	gi 1001453	hypothetical protein (Synecocystis sp.)	60	37	1014
33	1	26	469	gi 388109	regulatory protein (Enterococcus faecalis)	60	41	444
37	13	10414	9834	gi 1336656	Oxfl (Bacillus subtilis)	60	40	981
39	4	4364	4522	gi 4872	ORF 4 (Saccharomyces kluyveri)	60	47	159
41	1	2047	1025	gi 142822	D-alanine racemase cds (Bacillus subtilis)	60	39	1023
43	4	2474	3607	gi 468046	para-nitrobenzyl esterase (Bacillus subtilis)	60	40	1134
44	10	6756	7769	gi 414234	thif (Escherichia coli)	60	52	1014
45	10	8874	9074	gi 343949	var1(40.0) (Saccharomyces cerevisiae)	60	44	201
56	18	27842	26430	gi 468764	mecR gene product (Rhizobium meliloti)	60	35	1413
60	2	173	388	gi 1303864	YggQ (Bacillus subtilis)	60	33	216
63	2	357	1619	gi 467124	ureD; B229.C3.234 (Mycobacterium leprae)	60	43	1263
69	1	787	395	gi 1518853	OsfA (Salmonella typhimurium)	60	36	193
88	1	1	1188	gi 1480429	putative transcriptional regulator (Bacillus stearothermophilus)	60	30	1188
92	6	4735	3881	gi 349227	transmembrane protein (Escherichia coli)	60	37	855
92	7	5996	4923	gi 466613	nikB (Escherichia coli)	60	38	1076
93	1	949	476	gi 1510925	coenzyme F420-reducing hydrogenase, beta subunit (Methanococcus jannaschii)	60	27	474
96	6	7166	7178	gi 472715	accessory protein (Mycobacterium placentiae)	60	30	213
98	6	3212	4069	gi 467425	unknown (Bacillus subtilis)	60	42	858
102	10	7158	7430	gi 141092	acetolactate synthase small subunit (Bacillus subtilis) sp P37232 TLN-BACSU ACETOLACTATE SYNTHASE SMALL SUBUNIT (EC 1.3.18) (NIAS) (ACETOHYDROXY-ACID SYNTHASE SMALL SUBUNIT) (ALS1)	60	37	273
109	11	9127	10515	gi 1255259	o-succinylbenzoyl acid (OSB) CoA ligase (Staphylococcus aureus)	60	28	1389
109	12	110499	11656	gi 141954	beta-ketothiolase (Alicapigenes eutrophus)	60	41	1158
119	2	4630	3134	gi 1524280	unknown (Mycobacterium tuberculosis)	60	45	1497

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
121	9	6957	7646	gi 1107529	ceuC gene product [Campylobacter coli]	60	35	690
140	7	7704	6013	gi 146547	kdpA [Escherichia coli]	60	45	1692
145	1	2	703	gi 1460077	unknown [Mycobacterium tuberculosis]	60	23	702
150	3	2809	2216	gi 1146230	putative [Bacillus subtilis]	60	40	594
157	2	1389	961	gi 1103975	YqjX [Bacillus subtilis]	60	30	429
158	5	5125	4769	gi 1449288	unknown [Mycobacterium tuberculosis]	60	36	357
159	1	511	257	gi 580932	murD gene product [Bacillus subtilis]	60	43	255
160	1	159	1187	gi 1204532	hypothetical protein (GB:L19201_29) [Haemophilus influenzae]	60	34	1029
161	14	8249	7866	gi 1496003	ORF3; PepY; putative oligopeptidase based on homology with Lactococcus lactis pepF (GenBank Accession Number Z32522) [Caldicellulosiruptor saccharolyticus]	60	34	384
172	3	1331	2110	gi 485280	28.2 kDa protein [Streptococcus pneumoniae]	60	33	780
173	2	4082	2460	gi 1524397	glycine betaine transporter Opd [Bacillus subtilis]	60	41	1623
173	1	5063	4953	gi 1100737	NADP dependent isonitroreductase [Sus scrofa]	60	44	1011
198	1	3	985	gi 413943	ipa-19d gene product [Bacillus subtilis]	60	42	993
201	4	3641	4573	ep P37028 YADT_	HYPOTHETICAL 29.4 KD PROTEIN IN HEM-PPS INTERGENIC REGION PRECURSOR.	60	37	933
203	1	3269	2415	gi 927798	D9719.34p; CAT; 0.14 [Saccharomyces cerevisiae]	60	43	835
206	9	12234	12515	ep P37347 YECO_	HYPOTHETICAL 71.4 KD PROTEIN IN ASP5 5'-REGION.	60	47	282
212	4	1213	1410	gi 132711	hemagglutinin-neuraminidase fusion protein [human parainfluenza virus 3]	60	34	198
214	1	65	1153	gi 1204366	hypothetical protein (GB:U14003_130) [Haemophilus influenzae]	60	36	1089
237	1	2	937	gi 149377	Hsd [Lactococcus lactis]	60	40	936
241	6	5696	4998	gi 1046160	hypothetical protein (GB:U00021_5) [Mycoplasma genitalium]	60	37	699
260	6	5919	6485	gi 1431950	similar to a B.subtilis gene (GB: BACHEMY_5) [Clostridium acetabulum]	60	35	567
264	1	2432	1218	gi 1397526	clumping factor [Staphylococcus aureus]	60	53	1215
267	1	3	1409	gi 148316	NAH-antiprotein protein [Enterococcus faecalis]	60	27	1407
275	3	3804	4595	pir P36889 P368	leuD 3'-region hypothetical protein - Lactococcus lactis subsp. lactis (strain IL1403)	60	35	792
291	3	860	1198	gi 1208889	codon for by C. elegans cDNA yk130e12.5; contains C2H2-type zinc fingers [Caenorhabditis elegans]	60	33	339

TABLE 2

E. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
307	6	3421	3176	gi 1070014	protein-dependent [Bacillus subtilis]	60	36	246
316	8	4957	5823	gi 413952	lipa-28d gene product [Bacillus subtilis]	60	41	867
328	4	2596	3484	gi 1204484	membrane-associated component, branched amino acid transport system [Haemophilus influenzae]	60	39	489
332	5	4887	4363	gi 1205449	colicin V production protein (pur regulon) [Haemophilus influenzae]	60	37	525
357	1	1062	532	gi 487842	single-stranded DNA-specific exonuclease [Escherichia coli]	60	41	531
375	2	56	362	gi 1057	adenylyl cyclase gene product [Saccharomyces kluyveri] rJQ145[OYBYK adenylate cyclase [EC 4.6.1.1] - yeast ccharomyces kluyveri]	60	47	267
397	1	66	416	gi 709599	glucarate dehydratase [Bacillus subtilis]	60	37	351
409	1	2	163	gi 499700	glycogen phosphorylase [Saccharomyces cerevisiae]	60	35	162
453	4	914	1237	gi 1196899	unknown protein [Staphylococcus aureus]	60	36	324
453	7	3838	3620	gi 12222 YCF1_	HYPOTHEICAL 226 KD PROTEIN (ORF 1901)	60	31	219
470	2	622	945	gi 530782 S307	integrin homolog - yeast [Saccharomyces cerevisiae]	60	31	324
500	1	118	606	gi 467407	unknown [Bacillus subtilis]	60	36	489
503	3	752	982	gi 167835	myosin heavy chain [Dictyostelium discoideum]	60	34	231
505	4	2238	3563	gi 1510732	NADH oxidase [Methanococcus jannaschii]	60	26	1326
523	1	3	1043	gi 143331	alkaline phosphatase regulatory protein (Bacillus subtilis) p1r A27650 A27650 regulatory protein phoR - Bacillus subtilis ap P21545 PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN INOR [EC 2.7.3.-]	60	43	1041
543	1	1	465	gi 1511103	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	60	40	465
545	1	1	726	gi 1498192	putative [Pseudomonas aeruginosa]	60	40	726
556	1	2	1054	gi 1477402	tox gene product [Bordetella pertussis]	60	42	1053
578	1	974	489	gi 1205129	H. influenzae predicted coding region H10882 [Haemophilus influenzae]	60	42	486
594	1	1	624	gi 1212755	adenylyl cyclase [Aeromonas hydrophila]	60	45	624
604	1	3	530	gi 145925	fecB [Escherichia coli]	60	42	528
620	1	926	465	gi 1205483	bicyclicomycin resistance protein [Haemophilus influenzae]	60	33	462
630	2	871	1122	gi 1486242	unknown [Bacillus subtilis]	60	41	252
645	2	574	425	gi 1205136	serine hydroxymethyltransferase [serine methylase] [Haemophilus influenzae]	60	28	150

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
684	1	1082	843	gi 1205538	hypothetical protein (GB:U14003.302) (Haemophilus influenzae)	60	39	240
786	1	967	485	gi 1402944	orf8M1 gene product (Bacillus subtilis)	60	46	483
844	1	588	346	gi 790943	urea amidolyase (Bacillus subtilis)	60	40	243
851	1	1	726	gi 159661	ORP reductase (Ascaris lumbricoides)	60	41	726
871	1	1746	874	gi 1001493	hypothetical protein (Synchocystis sp.)	60	39	873
896	1	1558	839	gi 604926	NADH dehydrogenase, subunit 5 (Schizosaccharomyces pombe) sp P50366 NU5M_SCHCO NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 6.5.3)	60	39	720
908	2	448	753	gi 662880	novel hemolytic factor (Bacillus cereus)	60	31	306
979	1	2	595	gi 1429255	putative; orf1 (Bacillus subtilis)	60	30	594
1078	1	669	502	gi 581055	inner membrane copper tolerance protein (Escherichia coli) gi 871029 disulphide isomerase like protein (Escherichia coli) pIr S47295 S47295 inner membrane copper tolerance protein - Escherichia coli	60	40	168
1112	1	1150	620	gi 407885	ORF3 (Streptomyces griseus)	60	34	531
1135	1	484	275	gi 1171407	Vps8p (Saccharomyces cerevisiae)	60	36	210
1146	1	17	562	gi 123981	hypothetical protein (Bacillus subtilis)	60	36	546
1291	1	716	360	plr S57530 S575	carboxyl esterase - Acinetobacter calcoaceticus	60	30	357
1332	1	336	169	gi 1222056	aminotransferase (Haemophilus influenzae)	60	44	168
1429	1	3	146	gi 1205619	ferritin like protein (Haemophilus influenzae)	60	39	144
1722	1	570	286	gi 240052	dihydroflavonol-4-reductase, DFR (Hordeum vulgare-barley, cv. Gula, eptide, 354 aa)	60	36	285
2150	1	385	200	gi 497626	ORF 1 (Plasmid pAQ1)	60	20	186
2936	1	519	310	gi 508981	prephenate dehydratase (Bacillus subtilis)	60	48	210
3027	1	568	302	gi 1146199	putative (Bacillus subtilis)	60	37	267
3084	1	20	208	gi 1407784	orf-1; novel antigen (Staphylococcus aureus)	60	51	189
3155	1	2	226	gi 1046097	cytadherence-accessory protein (Mycoplasma genitalium)	60	34	225
3603	1	368	186	gi 510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA dehydrogenase alpha-subunit (Rattus norvegicus)	60	42	183
3665	1	486	244	gi 151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas mevalonil) pIr A44756 A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	60	42	243
3747	1	3	146	gi 474192	lucC gene product (Escherichia coli)	60	36	144

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match junction (%)	match gene name	% sim	% ident	length (nt)
3912	1	3	335	gi 488695	novel antigen: orf-2 [Staphylococcus aureus]	60	44	333
4072	1	3	272	gi 405879	yleiH [Escherichia coli]	60	33	270
4134	1	510	352	gi 780656	chemoreceptor protein [Rhizobium leguminosarum bv. viciae] gi 780656 chemoreceptor protein [Rhizobium leguminosarum bv. viciae]	60	28	359
4207	2	677	402	gi 602031	similar to trimethylamine DH [Mycoplasma capricolum] pir S49950 S49950 Probable trimethylamine dehydrogenase (EC 5.99.7) - Mycoplasma capricolum (SQC3) (fragment)	60	41	276
4243	1	127	324	gi 899317	peptide synthetase module [Microcystis aeruginosa] pir S49111 S49111 Probable amino acid activating domain - Microcystis aeruginosa (fragment) (SUB 144-328)	60	42	198
4110	1	624	313	gi 306980	phoB [Bacillus subtilis]	60	28	312
4345	1	343	173	gi 510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA dehydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
4382	1	498	280	gi 47382	acyl-CoA dehydrogenase [Streptomyces purpurascens]	60	48	219
4474	1	53	223	gi 510108	mitochondrial long-chain enoyl-CoA hydratase/3-hydroxycyl-CoA dehydrogenase alpha-subunit [Rattus norvegicus]	60	42	171
23	4	4518	3523	gi 426846	Vlpa protein [Salmonella typhi]	59	39	996
33	2	707	1483	pir S48604 S486	hypothetical protein - Mycoplasma capricolum (SQC3) (fragment)	59	33	772
33	5	4653	5853	gi 6721	F5982.3 [Caenorhabditis elegans]	59	33	1203
17	2	3228	2299	gi 142833	ORF2 [Bacillus subtilis]	59	37	910
38	21	16784	16593	gi 912576	81P [Phaeodactylum tricornutum]	59	40	192
52	3	2648	2349	gi 536972	ORP-090a [Escherichia coli]	59	44	300
54	12	14181	13402	gi 483940	transcription regulator [Bacillus subtilis]	59	37	780
57	3	4397	3339	gi 508176	Gat-1-P-DH, NAD dependant [Escherichia coli]	59	40	1059
66	1	986	495	gi 1303901	Yqhr [Bacillus subtilis]	59	34	492
67	7	6552	7460	gi 912461	nlkC [Escherichia coli]	59	37	909
70	7	5383	6366	gi 1339822	PhoD precursor [Rhizobium meliloti]	59	46	984
78	1	1	1449	gi 971345	unknown, similar to E.coli cardiolipin synthase [Bacillus subtilis] sp P45860 YME_BACSU HYPOTHETICAL 58.2 PROTEIN IN NARI-ACDA INTERGENIC REGION.	59	39	1449
82	10	14329	15534	gi 490328	LORF P (unidentified)	59	44	1206

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
89	2	1602	958	gi 642801	unknown (Saccharomyces cerevisiae)	59	32	645
96	4	4940	5473	gi 1333802	protein of unknown function (Rhodobacter capsulatus)	59	33	534
98	1	2	820	gi 467421	similar to B. subtilis DnaK (Bacillus subtilis)	59	34	819
119	1	166	1557	gi 143122	ORF B; putative (Bacillus firmus)	59	36	1392
120	10	6214	6756	gi 15354	ORF 55.9 (Bacteriophage T4)	59	39	543
120	16	12476	13510	gi 1086575	Beta (Rhizobium maitioli)	59	44	1035
123	1	386	195	gi 984737	catalase (Campylobacter jejuni)	59	38	192
130	1	370	645	gi 1256634	25.8% identity over 120 aa with the Synenococcus sp. Mpev protein; putative (Bacillus subtilis)	59	31	276
131	4	5278	5712	gi 1510655	hypothetical protein (SP:P42297) (Methanococcus jannaschii)	59	39	435
164	1	3	509	gi 1001342	hypothetical protein (Synechocystis sp.)	59	41	507
164	4	1529	2821	gi 1205165	hypothetical protein (SP:P37784) (Haemophilus influenzae)	59	35	1293
164	19	19643	21376	gi 1001381	hypothetical protein (Synechocystis sp.)	59	34	1734
173	3	4727	3717	gi 1184121	auxin-induced protein (Vigna radiata)	59	50	1011
179	2	2218	1688	gi 143036	unidentified gene product (Bacillus subtilis)	59	33	531
195	12	12669	11503	gi 762778	NIFS gene product (Anaerobaculum azollae)	59	41	1167
201	5	4702	5670	gi 3510240	hemolysin (Methanococcus jannaschii)	59	32	969
201	7	5719	6315	gi 1511456	M. jannaschii predicted coding region M1437 (Methanococcus jannaschii)	59	34	597
209	1	102	461	gi 1204666	hypothetical protein (CB:X7324_53) (Haemophilus influenzae)	59	42	360
214	3	1050	2234	gi 551531	2-nitropropane dioxygenase (Williopsis saturnus)	59	36	1185
214	5	3393	4135	gi 1303709	YrkJ (Bacillus subtilis)	59	32	843
217	2	1381	2167	gi 290489	dcp (CG Site No. 18430) (Escherichia coli)	59	44	1215
237	5	3078	3785	gi 149382	MISA (Lactococcus lactis)	59	38	708
251	2	376	960	gi 3303791	YqjJ (Bacillus subtilis)	59	34	585
286	1	1621	812	gi 146551	transmembrane protein (tdpD) (Escherichia coli)	59	31	810
316	5	4978	3860	gi 405879	yeiK (Escherichia coli)	59	32	1119
370	3	600	761	gi 1303794	YqeM (Bacillus subtilis)	59	35	162

TABLE 2

S. aureus - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
382	1	1009	506	gi 547513	orf3 (Haemophilus influenzae)	59	34	504
391	3	1620	1273	gi 152901	ORF 3 (Spirochaete aurantia)	59	37	348
406	3	2805	1705	gi 709992	hypothetical protein (Bacillus subtilis)	59	34	1101
426	5	3802	3245	gi 1204610	iron(III) dictrate transport ATP-binding protein PCEC (Haemophilus influenzae)	59	36	558
439	2	1513	1148	gi 1064809	homologous to sp.NTNA_ECOLI (Bacillus subtilis)	59	42	366
460	2	708	1301	gi 466882	ppsl; B1496_C2_189 (Mycobacterium leprae)	59	37	594
461	4	2212	3135	gi 1498295	homoserine kinase homolog (Streptococcus pneumoniae)	59	37	924
473	1	2929	1607	gi 147989	trigger factor (Escherichia coli)	59	40	1223
480	8	5862	6110	gi 1205311	(3R)-hydroxymyristoyl acyl carrier protein dehydrase (Haemophilus influenzae)	59	40	249
521	1	14	1354	gi 1256201256	staphylocoagulase - Staphylococcus aureus (fragment)	59	32	1341
534	4	2994	4073	gi 153746	mammitol-phosphate dehydrogenase (Streptococcus mutans) pIr-C44798[C44798]	59	36	1080
535	1	1	954	gi 1469939	mammitol-phosphate dehydrogenase M1D - treptococcus mutans	59	33	954
551	3	2836	3186	gi 1204511	group B oligopeptidase PepB (Streptococcus agalactiae)	59	45	351
573	2	449	940	gi 386681	bacterioferritin comigratory protein (Haemophilus influenzae)	59	36	492
650	1	5	748	gi 396400	ORF VAL022 (Saccharomyces cerevisiae)	59	30	744
664	1	566	285	gi 1262748	similar to eukaryotic Na <sup>+</sup> /H <sup>+</sup> exchanger (Escherichia coli) sp P32703 VJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS MTERGENIC REGION (O549)	59	33	282
670	1	3	455	gi 1122758	LuKP-PV like component (Staphylococcus aureus)	59	42	453
674	3	543	929	gi 293033	unknown (Bacillus subtilis)	59	46	387
758	1	349	176	gi 1500472	integrase (Bacteriophage phi-LC3)	59	37	174
771	2	2270	1461	gi 522150	M. jannaschii predicted coding region MJ1577 (Methanococcus jannaschii)	59	44	810
825	1	2191	1097	gi 397526	bromoperoxidase BPO-A1 (Streptomyces aureofaciens) sp P33912 BPAI_STRAU NON-HAEM BROMOPEROXIDASE BPO-A1 (EC 1.11.1.-) BROMIDE PEROXIDASE (BPO1)	59	47	1095
1052	2	1094	723	gi 289262	clumping factor (Staphylococcus aureus)	59	36	372
1152	1	373	188	gi 1276668	comE ORF3 (Bacillus subtilis)	59	37	186
					ORF218 gene product (Porphyra purpurea)			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1198	1	492	247	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	59	26	246
1441	1	468	235	gi 1045942	glycyl-tRNA synthetase [Mycoplasma genitalium]	59	37	234
2103	1	1	186	gi 459250	triacylglycerol lipase [Galactomyces geotrichum]	59	33	186
2205	1	793	398	gi 1303794	Yqem [Bacillus subtilis]	59	38	396
2578	1	484	284	gi 258003	insulin-like growth factor binding protein complex acid-labile subunit [rat, liver, Peptide, 601 aa]	59	48	201
2967	2	145	348	gi 1212730	Yqhk [Bacillus subtilis]	59	44	204
3012	1	3	248	gi 773571	neurofilament protein NF70 [Helix asperae]	59	31	246
3544	1	3	401	gi 1055218	crotonase [Clostridium acetobutylicum]	59	42	399
3548	1	3	401	gi 1055218	crotonase [Clostridium acetobutylicum]	59	42	399
3580	1	698	351	gi 1055218	crotonase [Clostridium acetobutylicum]	59	42	348
3720	1	722	363	gi 1408494	homologous to penicillin acylase [Bacillus subtilis]	59	36	360
4171	1	3	296	gi 1055218	crotonase [Clostridium acetobutylicum]	59	42	294
4305	1	618	310	gi 1524193	unknown [Mycobacterium tuberculosis]	59	39	309
18	1	1242	622	gi 146913	N-acetylglucosamine transport protein [Escherichia coli] pir b29895 WQBC2N phosphotransferase system enzyme II [EC 7.1.69], N-acetylglucosamine-specific - Escherichia coli sp P09322 PRAA_ECOLI PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC IIABC COMPONENT [ETIA]	58	43	621
20	7	7020	5845	gi 50502	collagen alpha chain precursor (AA -27 to 1127) [Mus musculus]	58	50	1176
21	5	3234	1626	gi 1034860	phosphoribosyl anthranilate isomerase [Thermotoga maritima]	58	32	393
23	2	2841	1669	gi 1276860	EpsG [Streptococcus thermophilus]	58	29	1173
23	10	9301	8090	pir A1133 A311	diaminopimelate decarboxylase [EC 4.1.1.20] - Pseudomonas aeruginosa	58	37	1212
38	29	22555	22884	gi 973249	vestitone reductase [Medicago sativa]	58	37	330
44	1	2	406	gi 289272	ferrichrome-binding protein [Bacillus subtilis]	58	33	405
45	1	1	552	gi 29464	embryonic myosin heavy chain (1085 AA) [Homo sapiens] Ir S12460 S12460 myosin beta heavy chain - human	58	33	552
55	2	759	538	gi 158852	glucose regulated protein [Schinococcus multilocularis]	58	32	222
62	13	8493	8068	gi 973353	kinase-associated protein B [Bacillus subtilis]	58	35	426
63	3	1553	1717	gi 164926	[Arabidopsis thaliana unidentified mRNA sequence, complete cds.], ene product [Arabidopsis thaliana]	58	35	165

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
67	13	12017	11229	gi11228083	[NADH dehydrogenase subunit 2 [Chorthippus parallelus]	58	41	789
96	8	8208	9167	gi1709992	[hypothetical protein [Bacillus subtilis]	58	42	960
107	2	2065	1364	gi1806327	[Escherichia coli hcpA gene for A protein similar to yeast PRP16 and RP22 [Escherichia coli]	58	37	702
112	7	4519	5613	gi1155588	[glucose-fructose oxidoreductase [Zymomonas mobilis] pir[A42289/A42289 glucose-fructose oxidoreductase (EC 1.1.-.-) reductase - Zymomonas mobilis]	58	38	1095
114	6	7118	6503	gi1137843	[unknown [Bacillus subtilis]	58	38	816
143	2	2261	1395	pir[A15605/A456	[mature-parasite-infected erythrocyte surface antigen HESA - Plasmodium falciparum]	58	31	867
151	2	717	950	gi11370261	[unknown [Mycobacterium tuberculosis]	58	31	234
154	6	6015	4627	gi11209277	[pCTH01 gene product [Chlamydia trachomatis]	58	41	1189
154	16	114281	113541	gi1146613	[DNA ligase (EC 6.5.1.2) [Escherichia coli]	58	33	741
155	3	2269	1892	gi11303917	[YqjB [Bacillus subtilis]	58	34	378
174	1	1056	529	gi1904198	[hypothetical protein [Bacillus subtilis]	58	26	528
189	4	1513	1769	gi1467383	[DNA binding protein (probable) [Bacillus subtilis]	58	25	237
201	3	2669	3307	gi11511453	[endonuclease III [Methanococcus jannaschii]	58	34	639
208	1	2	238	gi11276729	[phycobilisome linker polypeptide [Porphyra purpurea]	58	29	237
220	11	114575	11058	gi1397526	[clumping factor [Staphylococcus aureus]	58	51	1518
231	3	1629	1474	gi11002520	[HutS [Bacillus subtilis]	58	45	156
233	6	4201	3497	gi11463023	[No definition line found [Caenorhabditis elegans]	58	39	705
243	10	9303	10082	gi1537207	[ORF_1277 [Escherichia coli]	58	32	780
257	1	331	1143	gi11340128	[ORF1 [Staphylococcus aureus]	58	44	813
302	2	460	801	gi140176	[ORF X [Bacillus subtilis]	58	34	342
307	11	6986	6127	gi11303842	[YqjU [Bacillus subtilis]	58	30	858
321	3	1916	2747	gi11219996	[hypothetical protein [Bacillus subtilis]	58	41	834
342	4	2724	3497	gi1454838	[ORF 6; putative [Pseudomonas aeruginosa]	58	41	774
348	1	1	663	gi1467478	[unknown [Bacillus subtilis]	58	36	663
401	2	384	605	gi1143407	[para-aminobenzoic acid synthase, component 1 (pab) [Bacillus subtilis]	58	53	222

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
437	1	325	1554	gi1303866	Yqo5 [Bacillus subtilis]	58	35	1230
445	1	105	1482	gi1581583	protein A [Staphylococcus aureus]	58	32	1338
453	3	789	965	gi11009455	unknown [Schizosaccharomyces pombe]	58	34	177
453	5	2748	2047	gi1537214	yjiC gene product [Escherichia coli]	58	40	702
479	2	731	1444	gi1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	58	36	714
490	1	909	547	gi1580920	rodd (graA) polypeptide (AA 1-473) [Bacillus subtilis] p1rS06048 [S06048] probable rodd protein [Bacillus subtilis] sp1p1284 [PAGE_BACSU PROBABLE POLY(DL-CYCLO-PROPHATE) ALPHA-GLUCOSYLTRANSFERASE (EC 2.4.1.32) (TECHOTIC ACID BIOSYNTHESIS PROTEIN E)]	58	36	363
517	1	1	1164	sp1p47264 [V018_	HYPOTHETICAL HELICASE M0018.	58	30	1164
517	6	4182	4544	gi1453432	orf268 gene product [Hycoplasma hominis]	58	29	363
546	3	2802	4019	gi1866052	restriction modification system S subunit [Spiroplasma citri] gi1866052 restriction modification system S subunit [Spiroplasma citri]	58	37	1218
562	1	3	179	gi143831	infS protein (AA 1-400) [Klebsiella pneumoniae]	58	34	177
600	2	1347	1156	gi1181839	unknown [Pseudomonas aeruginosa]	58	48	192
604	2	1231	1001	gi11001353	hypothetical protein [Synecocystis sp.]	58	41	231
619	1	1	504	gi1903748	integral membrane protein [Homo sapiens]	58	43	504
625	1	2	364	gi11208474	hypothetical protein [Synecocystis sp.]	58	43	363
635	1	1492	755	gi11510995	transaldolase [Methanococcus jannaschii]	58	41	738
645	1	1	846	gi1677882	ileal sodium-dependent bile acid transporter [Rattus norvegicus] gi1677882 ileal sodium-dependent bile acid transporter [Rattus norvegicus]	58	33	846
645	3	906	1556	gi11239999	hypothetical protein [Bacillus subtilis]	58	41	651
665	1	771	532	gi11204262	hypothetical protein (CB:L10328_61) [Haemophilus influenzae]	58	39	240
674	1	635	327	gi1498817	ORF8: homologous to small subunit of phage terminase [Bacillus subtilis]	58	39	309
675	2	1312	806	gi142181	osmC gene product [Escherichia coli]	58	28	507
745	1	618	310	gi11205432	coenzyme PQ synthesis protein III (pqgIII) [Haemophilus influenzae]	58	32	309
799	2	242	1174	gi11204669	collagenase [Haemophilus influenzae]	58	36	933
800	2	1096	614	gi1171963	tRNA isopentenyl transferase [Saccharomyces cerevisiae] sp1p07884 [MOD5: YEAST tRNA ISOPENTENYLTRANSFERASE (EC 2.5.1.8) ISOPENTENYL-DIPHOSPHATE: tRNA ISOPENTENYLTRANSFERASE] (tpp transferase) (tppT).	58	37	483

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
854	1	1108	605	gi 466778	lysine specific peptidase [Escherichia coli]	58	44	504
885	1	481	242	gi 861199	protoporphyrin IX Mg-chelatase subunit precursor [Hordeum vulgare]	58	33	240
891	1	3	527	gi 1293660	AbsA2 [Streptomyces coelicolor]	58	31	525
942	1	931	467	gi 405567	traH [plasmid pSK41]	58	30	465
1002	1	952	521	gi 577649	preLJUH [Staphylococcus aureus]	58	34	432
1438	1	1	261	gi 581558	Isolucyl tRNA synthetase [Staphylococcus aureus] sp P41368 SVIP_STAAU ISOLEUCYL-TRNA SYNTHETASE, MUPIROICIN RESISTANT EC 6.1.1.5 (ISOLEUCINE--TRNA LIGASE) (ILERS) (MUPIROICIN RESISTANCE ROTENIN)	58	30	261
1442	1	2	463	gi 971394	similar to Acc.No. D26185 [Escherichia coli]	58	34	462
1873	1	480	241	gi 1339951	small subunit of NADH-dependent glutamate synthase [Plectonaea boryanum]	58	38	240
1876	1	3	158	gi 529216	No definition line found [Caenorhabditis elegans] sp P4650 YUX7_CAEEL HYPOTHETICAL 7.3 KD PROTEIN F23F12.7 IN CHROMOSOME III	58	33	156
1989	1	108	401	gi 1405458	YneR [Bacillus subtilis]	58	29	294
2109	1	3	401	gi 1001801	hypothetical protein [Synecocystis sp.]	58	31	399
2473	1	288	145	gi 510140	lipoendopeptidase F [Lactococcus lactis]	58	38	144
2523	1	452	228	gi 644873	catabolic dehydroquinase dehydratase [Acinetobacter calcoaceticus]	58	37	225
3041	1	2	211	gi 1205367	oligopeptide transport ATP-binding protein [Haemophilus influenzae]	58	39	210
3094	1	3	263	gi 1185288	isochorismate synthase [Bacillus subtilis]	58	38	261
3706	1	3	383	gi 456614	mevalonate kinase [Arabidopsis thaliana]	58	48	381
3854	1	1	402	gi 808869	human gcp372 [Homo sapiens]	58	32	402
4082	1	51	224	gi 504551	ribulose-1,5 biphosphate carboxylase large subunit -methyltransferase [Pisum sativum]	58	37	174
4278	1	3	206	gi 180189	cerebellar-degeneration-related antigen (CDR14) [Homo sapiens] gi 182737 cerebellar degeneration-associated protein [Homo sapiens] p1c A29770 A29770 cerebellar degeneration-related protein - human	58	37	204
19	7	7818	7363	gi 1001516	hypothetical protein [Synecocystis sp.]	57	31	456
23	11	9663	8872	gi 605066	ORF_F256 [Escherichia coli]	57	29	792
31	1	4801	2402	gi 153146	ORF3 [Streptomyces coelicolor]	57	32	2100
38	14	11611	10796	gi 144859	ORF B [Clostridium perfringens]	57	31	816
46	14	12063	113046	gi 1001319	hypothetical protein [Synecocystis sp.]	57	25	984

TABLE 2

S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
51	3	1411	1187	gi 131856 b338	hypothetical 80K protein - <i>Bacillus aptharicus</i>	57	38	225
54	1	1	453	gi 1684950	staphylococcal accessory regulator A ( <i>Staphylococcus aureus</i> )	57	31	453
75	1	3	239	gi 1000470	C2787.7 ( <i>Caenorhabditis elegans</i> )	57	42	237
92	5	3855	3061	gi 143607	sporulation protein ( <i>Bacillus subtilis</i> )	57	35	785
96	3	4006	4773	gi 144297	acetyl esterase (XycC) ( <i>Caldocellum saccharolyticum</i> ) pir B37202 B37202	57	34	768
107	3	1480	2076	gi 1460955	acetyl esterase (EC 3.1.1.6) (XycC) - <i>Caldocellum saccharolyticum</i>	57	42	597
109	8	5340	5933	gi 1438846	TagS ( <i>Vibrio cholerae</i> )	57	41	594
112	9	6679	7701	gi 1486250	Unknown ( <i>Bacillus subtilis</i> )	57	33	1023
114	4	6384	8108	gi 871456	Unknown ( <i>Bacillus subtilis</i> )	57	37	2277
126	2	430	1053	gi 288301	putative alpha subunit of formate dehydrogenase ( <i>Methanobacterium thermoautotrophicum</i> )	57	37	624
131	5	6537	8277	gi 1511160	ORF2 gene product ( <i>Bacillus megaterium</i> )	57	38	261
133	3	2668	2201	gi 1303912	M. jannaschii predicted coding region M31163 ( <i>Methanococcus jannaschii</i> )	57	40	468
133	4	3383	2784	gi 1221884	Yqhw ( <i>Bacillus subtilis</i> )	57	37	600
147	4	2164	1694	gi 467469	{urea?} amidolyase ( <i>Haemophilus influenzae</i> )	57	33	471
160	2	1293	1060	gi 558604	Unknown ( <i>Bacillus subtilis</i> )	57	28	234
163	8	5887	4764	gi 145580	chitin synthase 2 ( <i>Neurospora crassa</i> )	57	38	924
168	6	4336	5325	gi 39782	chitin synthase product ( <i>Escherichia coli</i> )	57	32	990
170	5	3297	3455	gi 603404	33kDa lipoprotein ( <i>Bacillus subtilis</i> )	57	37	159
221	6	8026	6809	gi 1136221	Ver16p ( <i>Baccharomyces cerevisiae</i> )	57	32	1218
228	3	1348	1791	gi 288969	carboxypeptidase ( <i>Sulfolobus solfataricus</i> )	57	32	444
263	4	4411	3686	gi 1185002	[fibronectin binding protein ( <i>Streptococcus dysgalactiae</i> ) pir S33850 S33850	57	42	726
276	1	494	255	gi 396380	[fibronectin-binding protein - <i>Streptococcus pyogenes</i> ]	57	40	240
283	2	335	1324	gi 773349	[dihydrodipicolinate reductase ( <i>Pseudomonas syringae</i> pv. <i>tabaci</i> )	57	32	990
297	1	469	236	gi 1334820	[No definition line found ( <i>Escherichia coli</i> )	57	46	234
342	3	1993	2805	gi 1204431	BirA protein ( <i>Bacillus subtilis</i> )	57	35	813
					[reading frame V ( <i>Cauliflower mosaic virus</i> )			
					[hypothetical protein (SP:P33644) ( <i>Haemophilus influenzae</i> )			

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
325	6	3140	3741	gi 305177	cell division protein [Bacillus subtilis]	57	26	402
433	6	3286	4011	gi 1524117	alpha-acetolactate decarboxylase [Lactococcus lactis]	57	40	726
470	3	903	1145	gi 104819	protein serine/threonine kinase [Toxoplasma gondii]	57	30	243
487	5	1391	1723	gi 507223	ORF1 [Bacillus stearothermophilus]	57	28	333
498	1	274	852	gi 1334549	MDM-ubiquitin oxidoreductase subunit A [Podospore anserina]	57	34	579
503	1	343	173	gi 1502283	organic cation transporter OCT2 [Rattus norvegicus]	57	30	171
505	2	1619	1284	gi 466884	B1496_C2_194 [Mycobacterium leprae]	57	40	336
519	2	1182	2549	gi 1303707	YrkH [Bacillus subtilis]	57	34	1368
522	2	3234	1945	gi 1064809	homologue to spiHRA_ECO1 [Bacillus subtilis]	57	36	1290
538	2	909	1415	gi 153179	phosphorothiohyacin n-acetyltransferase [Streptomyces coelicolor]	57	40	507
547	1	968	486	gi 467340	unknown [Bacillus subtilis]	57	50	483
599	1	1062	532	gi 120692 TYRA_	PREPHEMATE DEHYDROGENASE (EC 1.3.1.12) (PDIH)	57	41	531
620	2	757	572	gi 1107894	unknown [Schizosaccharomyces pombe]	57	38	186
622	2	1600	1130	gi 173028	chlorodioxin II [Saccharomyces cerevisiae]	57	39	471
625	2	362	1114	gi 1262366	hypothetical protein [Mycobacterium leprae]	57	34	753
680	1	1	204	gi 143544	RNA polymerase sigma-30 factor [Bacillus subtilis] pir A28625 A28625 transcription initiation factor sigma H - ctillus subtilis	57	30	204
690	1	3	629	gi 466520	pocR [Salmonella typhimurium]	57	29	627
696	1	2	433	gi 413972	ipa-48r gene product [Bacillus subtilis]	57	33	432
704	1	36	638	gi 149931	M. jannaschii predicted coding region MJ1083 [Methanococcus jannaschii]	57	36	603
732	1	2316	1621	gi 1418999	orf4 [Lactobacillus sake]	57	37	696
746	1	451	227	gi 392973	han3 [Apolyta californica]	57	42	225
757	1	20	466	gi 43979	L. curvatus small cryptic plasmid gene for rep protein [Lactobacillus rvaus]	57	45	447
862	1	2	295	gi 1303827	YqfI [Bacillus subtilis]	57	21	294
1049	1	907	455	gi 1510108	ORF-1 [Agrobacterium tumefaciens]	57	35	453
1117	1	1387	695	gi 896286	INH2 terminus uncertain [Leishmania tarentolae]	57	28	693

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
1136	1	2	322	gi1303853	YqgF (Bacillus subtilis)	57	38	321
1144	2	1033	611	gi1310083	voltage-activated calcium channel alpha-1 subunit (Rattus norvegicus)	57	46	423
1172	1	1472	738	gi1511146	M. jannaschii predicted coding region M1143 (Methanococcus jannaschii)	57	28	735
1500	2	746	558	gi142780	putative membrane protein; putative (Bacillus subtilis)	57	35	189
1676	1	659	399	gi1313777	lucA11 permease (Escherichia coli)	57	31	261
2481	1	2	400	gi1237015	ORP4 (Bacillus subtilis)	57	23	399
3099	1	3	230	gi1204540	isochloramate synthase (Haemophilus influenzae)	57	39	228
3122	1	360	181	gi1882472	ORP_0466 (Escherichia coli)	57	40	180
3560	1	2	361	gi153490	tetracycline C resistance and export protein (Streptomyces lauroscens)	57	37	360
3850	1	856	434	gi115588	glucose-fructose oxidoreductase (Zymomonas mobilis) pir[A42289/A42289 glucose-fructose oxidoreductase (EC 1.1.1.-) recursor - Zymomonas mobilis]	57	40	423
3931	1	704	354	gi1413953	ipa-29d gene product (Bacillus subtilis)	57	36	351
3993	1	1	384	gi151259	HMG-CoA reductase (EC 1.1.1.88) (Pseudomonas nevalonii) pir[A44756/A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.]	57	39	384
4065	1	793	398	pir JV0037 RDEC	nitrate reductase (EC 1.7.99.4) alpha chain - Escherichia coli	57	31	396
4100	1	596	300	gi1086633	T06C10.5 gene product (Caenorhabditis elegans)	57	47	297
4163	1	571	287	gi121512	patatin (Solanum tuberosum)	57	50	285
4267	2	631	335	gi1000165	Spotting (Bacillus subtilis)	57	38	297
4358	1	3	302	gi1298032	EP (Streptococcus suis)	57	32	300
4389	2	108	290	gi1405894	1-phosphofructokinase (Escherichia coli)	57	37	183
4399	1	2	232	gi11483603	pristinamycin I synthase 1 (Streptomyces pristinaespiralis)	57	35	231
4481	1	572	288	gi1405879	yeiH (Escherichia coli)	57	44	285
4486	1	512	258	gi1515938	glutamate synthase (ferredoxin) (Synecocystis sp.) pir[S46957/S46957 glutamate synthase (ferredoxin) (EC 1.4.7.1) - Synecocystis sp.]	57	42	255
4510	1	481	242	gi1205301	leukotoxin secretion ATP-binding protein (Haemophilus influenzae)	57	38	240
4617	1	468	256	gi1511222	restriction modification enzyme, subunit M1 (Methanococcus jannaschii)	57	35	213
4	11	12201	11524	gi149204	histidine utilization repressor G (Klebsiella aerogenes) pir[A36730/A36730 hutG protein - Klebsiella pneumoniae (fragment) sp P19452 HUTG_KLEAE FORMINOGLOUTAMASE (EC 3.5.3.8) FORMINOGLOUTAMATE HYDROLASE (HISTIDINE UTILIZATION PROTEIN G) FRAGMENT)	56	31	678

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	8	4248	5177	gi 1322222	RACH1 (Homo sapiens)	56	33	930
38	28	21179	22264	gi 1480705	lipote-protein ligase (Mycoplasma capricolum)	56	34	1086
44	3	1861	2421	gi 490320	Y gene product (unidentified)	56	31	561
44	15	10103	10606	gi 1205099	hypothetical protein (CB:U19201_1) (Haemophilus influenzae)	56	39	504
50	6	4820	5161	gi 209931	fiber protein (Human adenovirus type 5)	56	48	342
53	4	2076	2972	gi 623476	transcriptional activator (Providencia stuartii) sp P43463 AARP_PROST TRANSCRIPTIONAL ACTIVATOR AARP.	56	30	897
67	6	5656	6594	gi 466613	nikB (Escherichia coli)	56	32	939
89	3	2364	1810	gi 482922	protein with homology to pail repressor of B. subtilis (Lactobacillus elbrueckii)	56	39	555
96	1	203	913	gi 145594	cAMP receptor protein (crp) (Escherichia coli)	56	35	711
109	21	18250	17846	gi 1204367	hypothetical protein (CB:U14003_278) (Haemophilus influenzae)	56	27	405
112	8	5611	6678	gi 155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pfr A42289 A42289 glucose-fructose oxidoreductase (EC 1.1.1.-) recursor - Zymomonas mobilis	56	40	1068
131	3	6404	5100	gi 619724	MgtE (Bacillus firmus)	56	30	1305
138	2	65	232	gi 413948	lpa-24d gene product (Bacillus subtilis)	56	31	168
138	4	823	1521	gi 580868	lpa-22f gene product (Bacillus subtilis)	56	31	699
146	2	740	447	gi 1046009	M. genitalium predicted coding region MG309 (Mycoplasma genitalium)	56	37	294
149	2	1639	1067	gi 945380	terminase small subunit (Bacteriophage LU-11)	56	35	573
163	1	2	223	gi 143947	glutamine synthetase (Bacteroides fragilis)	56	30	222
166	5	6745	6449	gi 405792	ORF154 (Pseudomonas putida)	56	26	297
187	1	31	393	gi 311237	H(+)-transporting ATP synthase (Zea mays)	56	30	363
190	1	2	373	gi 1109686	Prox (Bacillus subtilis)	56	35	372
191	8	11538	9943	gi 581070	acyl coenzyme A synthetase (Escherichia coli)	56	35	1596
195	1	1291	647	gi 1510242	collagenase (Methanococcus jannaschii)	56	34	645
230	3	2323	2072	gi 40363	heat shock protein (Clostridium acetobutylicum)	56	39	252
238	5	3383	3725	gi 1477533	sarA (Staphylococcus aureus)	56	31	393
270	2	813	1712	gi 765073	autolysin (Staphylococcus aureus)	56	41	900

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
290	1	3221	1632	gi 547513	[orf3] Haemophilus influenzae	56	34	1590
297	5	1140	1373	gi 3511556	[M. jannaschii predicted coding region MJ1561] Methanococcus jannaschii	56	40	234
321	2	2947	1799	gi 1001801	[hypothetical protein] Synecocystis sp.1	56	31	1149
359	2	1279	641	gi 46236	[noli gene product] Rhizobium meliloti	56	26	639
371	2	360	1823	gi 145304	[L-ribulokinase] Escherichia coli	56	39	1464
391	4	1762	2409	gi 1001634	[hypothetical protein] Synecocystis sp.1	56	34	648
402	1	380	192	gi 1438904	[5-HT4L receptor] Homo sapiens	56	48	189
416	4	2480	2109	gi 1408486	[HS74A gene product] Bacillus subtilis	56	31	372
424	3	1756	2334	gi 142471	[acetolactate decarboxylase] Bacillus subtilis	56	32	579
457	1	1907	1017	gi 1205194	[formamidopyrimidine-DNA glycosylase] Haemophilus influenzae	56	36	891
458	2	2423	1812	gi 15486	[terminase] Bacteriophage SPPI	56	17	612
504	2	2152	1283	gi 1142681	[lpp38] Pasteurella hemolytica	56	38	870
511	1	1	1284	gi 217049	[brnQ protein] Salmonella typhimurium	56	37	1244
604	3	1099	1701	gi 462109	[rim: 30S ribosomal protein S18 alanine acetyltransferase: 229_C1_170] Mycobacterium leprae	56	43	603
660	5	3547	3774	gi 1229106	[2K910.1] Caenorhabditis elegans	56	30	228
707	1	35	400	gi 153929	[NADPH-auxite reductase flavoprotein component] Salmonella typhimurium	56	38	366
709	2	1385	1095	gi 1510801	[hydrogenase accessory protein] Methanococcus jannaschii	56	38	291
718	1	1	495	gi 413948	[ipa-24d gene product] Bacillus subtilis	56	35	495
744	1	87	677	gi 928836	[repressor protein] Lactococcus lactis phage BK5-7	56	35	591
750	1	776	399	gi 1511513	[ABC transporter, probable ATP-binding subunit] Methanococcus jannaschii	56	33	378
755	1	3	407	gi 1205382	[cell division protein] Haemophilus influenzae	56	34	405
813	1	19	930	gi 1222161	[permease] Haemophilus influenzae	56	28	912
855	1	3	515	gi 1256621	[26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative] Bacillus subtilis	56	33	513
968	1	2	466	gi 547513	[orf3] Haemophilus influenzae	56	37	465
973	2	1049	732	gi 1866022	[MexR] Pseudomonas aeruginosa	56	31	318
1203	1	5	223	gi 184251	[HMG-1] Homo sapiens	56	34	219

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
1976	1	452	237	gi 19806	lysine-rich aspartic acid-rich protein (Plasmidium chabaudi) r[S22183]S22183 lysine/aspartic acid-rich protein - Plasmidium baudii	56	33	216
2161	1	2	400	gi 1237015	ORF4 (Bacillus subtilis)	56	27	399
2958	1	362	183	gi 466685	no definition line found (Escherichia coli)	56	26	180
2979	1	421	212	gi 1206354	spore germination and vegetative growth protein (Haemophilus influenzae)	56	40	210
2994	2	526	326	gi 836646	phosphoribosylformylino-praic ketolase (Rhodobacter phaeoideus)	56	29	201
3026	1	179	328	gi 143306	penicillin V amidase (Bacillus sphaericus)	56	30	150
3189	1	289	146	gi 1166604	similar to aldehyde dehydrogenase (Caenorhabditis elegans)	56	37	144
3770	1	63	401	gi 1129145	acetyl-CoA C-acyltransferase (Mongifera indicol)	56	43	339
4054	2	720	361	gi 1205355	[Na+/H+] antiporter (Haemophilus influenzae)	56	31	360
4145	1	1	324	gi 126095	long-chain acyl-CoA dehydrogenase (Mus musculus)	56	36	324
4200	1	505	254	gi 155588	glucose-fructose oxidoreductase (Zymomonas mobilis) pIR[A42289]A42289 glucose-fructose oxidoreductase (EC 1.1.1.-) recursor - Zymomonas mobilis	56	40	252
4273	1	675	355	gi 308861	GTP start codon (Lactococcus lactis)	56	33	321
1	3	4095	3436	gi 5341	Putative orf YCLX8c, len:192 (Saccharomyces cerevisiae) r[S53591]S53591 hypothetical protein - yeast (Saccharomyces evisiae)	55	25	660
11	12	9377	8505	gi 216773	haloacetate dehalogenase H-1 (Moraxella sp.)	55	32	873
12	4	5133	4534	gi 467337	unknown (Bacillus subtilis)	55	26	600
19	5	5404	5844	gi 1001719	hypothetical protein (Synecocystis sp.)	55	25	441
23	13	14087	12339	gi 474190	lucA gene product (Escherichia coli)	55	30	1749
32	7	5364	6888	gi 1340096	unknown (Mycobacterium tuberculosis)	55	37	1521
34	3	2589	1808	gi 1303968	YqjQ (Bacillus subtilis)	55	39	762
34	5	3960	3412	gi 1303962	YqjK (Bacillus subtilis)	55	33	549
36	1	1291	647	gi 406045	ORF_0118 (Escherichia coli)	55	27	645
36	6	6220	5243	gi 1001341	hypothetical protein (Synecocystis sp.)	55	31	978
47	3	3054	3821	gi 1001819	hypothetical protein (Synecocystis sp.)	55	21	768
49	1	2065	1127	gi 403373	glycerophosphoryl diester phosphodiesterase (Bacillus subtilis) pIR[S37251]S37251 glycerophosphoryl diester phosphodiesterase - acillus subtilis	55	36	939

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
67	11	8966	9565	gi 1151053	norA199 protein [Staphylococcus aureus]	55	23	600
75	3	881	1273	gi 12698	L-histidinol: NAD: oxidoreductase (EC 1.1.1.23) [aa 1-434] scherichia coli	55	33	393
82	9	11587	14194	gi 1136221	carboxypeptidase [Sulfolobus solfataricus]	55	35	1194
87	4	3517	4917	gi 1064812	function unknown [Bacillus subtilis]	55	26	1401
88	2	1172	1636	gi 982463	protein-Nip1-phosphohistidine-sugar phosphotransferase [Escherichia coli]	55	35	465
92	1	127	516	gi 11377832	unknown [Bacillus subtilis]	55	36	390
100	2	816	2035	gi 11370274	seaxanthin epoxidase [Nicotiana glauca]	55	36	1200
100	5	5137	4658	gi 396660	unknown open reading frame [Buchnera aphidicola]	55	29	480
108	3	4266	2986	gi 1499866	M. jannaschii predicted coding region M1024 [Methanococcus jannaschii]	55	31	1281
116	3	2616	1834	gi 1511367	formate dehydrogenase, alpha subunit [Methanococcus jannaschii]	55	29	783
144	3	1805	1476	gi 1100787	unknown [Saccharomyces cerevisiae]	55	35	310
165	5	6212	5508	gi 1045884	M. genitalium predicted coding region HQ199 [Mycoplasma genitalium]	55	27	705
189	5	2205	2576	gi 1142569	ATP synthase a subunit [Bacillus firmus]	55	35	372
191	6	9136	6857	gi 559411	B0272.3 [Caenorhabditis elegans]	55	39	2280
194	2	364	636	gi 1145768	K7 kinesin-like protein [Dictyostelium discoideum]	55	34	273
209	4	1335	1676	gi 473357	thi4 gene product [Schizosaccharomyces pombe]	55	35	342
211	2	1693	1145	gi 410130	ORFX6 [Bacillus subtilis]	55	37	549
213	2	644	1372	gi 633692	TrxA [Yersinia enterocolitica]	55	28	729
214	7	4144	5481	gi 1001793	hypothetical protein [Synecocystis sp.]	55	30	1338
221	7	11473	9197	gi 466520	pocR [Salmonella typhimurium]	55	32	2272
233	8	5908	4817	gi 11237063	unknown [Mycobacterium tuberculosis]	55	38	1092
236	4	1375	2340	gi 1146199	putative [Bacillus subtilis]	55	32	966
243	2	380	1885	gi 459907	mercuric reductase [Plasmid p1258]	55	29	1506
258	1	786	394	gi 455006	orf6 [Rhodococcus fasciens]	55	36	393
281	1	126	938	gi 1408493	homologous to SwissProt:YIDA_EC001 hypothetical protein [Bacillus subtilis]	55	35	813
316	3	1323	2102	gi 1486447	LuxA homologue [Rhizobium sp.]	55	30	780
326	5	2968	2744	gi 1296824	proline iminopeptidase [Lactobacillus helveticus]	55	36	235

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
351	2	2322	1429	gi1204820	hydrogen peroxide-inducible activator [Haemophilus influenzae]	55	24	894
353	4	2197	2412	gi1272475	chitin synthase [Enterococcus nidulans]	55	50	216
380	1	14	379	gi142554	ATP synthase 1 subunit [Bacillus megaterium]	55	37	366
383	1	462	232	gi149272	ferrichrome-binding protein [Bacillus subtilis]	55	36	231
386	1	3	936	gi1510251	DNA helicase, putative [Methanococcus jannaschii]	55	30	936
410	2	1208	1891	gi1205144	multidrug resistance protein [Haemophilus influenzae]	55	27	684
483	2	411	803	gi1413934	ipa-10r gene product [Bacillus subtilis]	55	26	423
529	3	1777	1433	gi1606150	ORF f309 [Escherichia coli]	55	33	345
555	1	1088	585	gi143407	para-aminobenzoic acid synthase, component 1 (pab) [Bacillus subtilis]	55	28	504
565	1	402	202	gi1223961	CDP-tylase epimerase [Yersinia pseudotuberculosis]	55	41	201
582	1	751	452	gi1256643	20.2% identity with NADH dehydrogenase of the Leishmania major mitochondrion; putative [Bacillus subtilis]	55	36	300
645	5	2260	2057	gi1210824	fusion protein F (bovine respiratory syncytial virus) p1rJ01481 VGNZ8A (fusion glycoprotein precursor - bovine respiratory syncytial virus tetra- A51908)	55	25	204
672	2	957	2216	gi1511333	M. jannaschii predicted coding region M31322 [Methanococcus jannaschii]	55	36	1280
710	1	955	479	gi1537007	ORF f379 [Escherichia coli]	55	30	477
737	1	1859	945	gi1536063	CG Site No. 16166 [Escherichia coli]	55	30	915
742	2	228	572	gi1304160	product unknown [Bacillus subtilis]	55	38	345
817	2	1211	903	gi11136289	histidine kinase A [Dictyostellium discoideum]	55	29	309
819	1	582	355	gi1558073	polymorphic antigen [Plasmodium falciparum]	55	22	228
832	2	1152	724	gi140367	ORF C [Clostridium acetobutylicum]	55	32	429
840	1	769	386	gi1203875	pseudouridylylase synthase 1 [Haemophilus influenzae]	55	39	384
1021	1	23	529	gi148563	beta-lactamase [Yersinia enterocolitica]	55	38	507
1026	1	60	335	gi147804	Opp C (AA1-301) [Salmonella typhimurium]	55	26	276
1325	1	1	282	gi1477533	isrA [Staphylococcus aureus]	55	29	282
1814	2	224	985	gi1046078	M. genitalium predicted coding region M3369 [Mycoplasma genitalium]	55	38	762
3254	1	427	254	gi1413968	ipa-44d gene product [Bacillus subtilis]	55	30	174

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	indicated gene name	% sim	% ident	length (nt)
3695	1	686	345	gi 216773	haloacetate dehalogenase H-1 (Moraxella sp.)	55	32	342
3721	1	1	312	gi 42029	ORF1 gene product (Escherichia coli)	55	31	312
3799	1	3	272	gi 42029	ORF1 gene product (Escherichia coli)	55	38	270
3889	1	22	423	gi 1129145	acetyl-CoA C-acyltransferase (Mangifera indica)	55	45	402
3916	1	2	385	gi 529754	speC (Streptococcus pyogenes)	55	38	384
3945	1	4	198	gi 476252	phage 1 (flagellin (Salmonella enterica))	55	36	195
4074	1	488	246	gi 42029	ORF1 gene product (Escherichia coli)	55	38	243
4184	1	2	343	gi 1524267	unknown (Mycobacterium tuberculosis)	55	28	342
4284	1	14	208	gi 1100774	ferredoxin-dependent glutamate synthase (Synecocystis sp.)	55	36	195
4457	2	644	378	gi 180189	cerebellar-degeneration-related antigen (CDR34) (Homo sapiens) cerebellar degeneration-associated protein (Homo sapiens) pif A29770 A29770 cerebellar degeneration-related protein - human	55	38	267
4514	1	2	244	gi 216773	haloacetate dehalogenase H-1 (Moraxella sp.)	55	32	243
4599	1	432	217	gi 1129145	acetyl-CoA C-acyltransferase (Mangifera indica)	55	42	216
4606	1	416	210	gi 386120	myosin alpha heavy chain (S2 subfragment) (rabbits, measeter, eptide Partial, 234 aa)	55	27	207
5	8	5348	4932	gi 536069	ORF YBL047c (Saccharomyces cerevisiae)	54	27	417
12	7	7166	6165	gi 1205504	homoserine acetyltransferase (Haemophilus influenzae)	54	30	1002
23	16	17086	15326	gi 474192	lucC gene product (Escherichia coli)	54	31	1761
35	1	2	979	gi 48034	small subunit of soluble hydrogenase (AA 1-184) (Synecococcus sp.) ir S06919 HYC55 soluble hydrogenase (EC 1.12.-.-) small chain - inchoococcus sp. (PCC 6716)	54	36	978
37	11	9417	8667	gi 537207	ORF_2277 (Escherichia coli)	54	38	771
37	12	8165	8132	gi 1160967	palmitoyl-protein thioesterase (Homo sapiens)	54	37	168
46	15	13025	13804	gi 438473	protein is hydrophobic, with homology to E. coli Prow; putative Bacillus subtilis	54	28	780
56	2	203	736	gi 1256139	YbbJ (Bacillus subtilis)	54	34	534
57	13	11117	10179	gi 1151248	inosine-uridine preferring nucleoside hydrolase (Citrobacter fasciculatus)	54	32	939
66	2	516	1133	gi 1335781	Cap (Drosophila melanogaster)	54	29	618
70	10	8116	8646	gi 1399823	PhoE (Rhizobium meliloti)	54	31	531

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match	accession	match	gene name	% sim	% ident	length (nt)
70	15	112556	11801	100	sp P02983 TCR_S	100	TETRAPEPTIDE RESISTANCE PROTEIN	54	29	756
87	5	4915	5706	100	gi 1064811	100	function unknown [Bacillus subtilis]	54	33	792
92	4	3005	2289	100	gi 1205366	100	oligopeptide transport ATP-binding protein [Haemophilus influenzae]	54	33	717
103	2	2596	1556	100	gi 1710495	100	protein kinase [Bacillus brevis]	54	33	1041
105	2	3585	2095	100	gi 143727	100	putative [Bacillus subtilis]	54	30	1491
112	4	2137	2732	100	gi 153724	100	MalC [Streptococcus pneumoniae]	54	41	396
127	2	1720	2493	100	gi 144297	100	acetyl esterase (XyxC) [Caldocellum saccharolyticum] pir B37202 B37202	54	34	774
138	5	1600	3306	100	gi 42473	100	pyruvate oxidase [Escherichia coli]	54	36	1707
152	2	525	1172	100	gi 1377834	100	unknown [Bacillus subtilis]	54	23	648
161	9	4831	5469	100	gi 903305	100	ORF3 [Bacillus subtilis]	54	28	639
161	13	6694	7251	100	gi 1511039	100	phosphate transport system regulatory protein [Methanococcus jannaschii]	54	32	558
164	6	3263	4543	100	gi 1204976	100	proyl-tRNA synthetase [Haemophilus influenzae]	54	34	1281
164	20	21602	22243	100	gi 143582	100	spolIIIEA protein [Bacillus subtilis]	54	32	642
171	6	5683	4230	100	gi 436965	100	[malA] gene producta [Bacillus stearothermophilus] pir S43914 S43914	54	37	1434
206	18	19208	19720	100	gi 1240016	100	hypothetical protein 1 - Bacillus teurothermophilus	54	38	513
218	2	1090	1905	100	gi 467378	100	unknown [Bacillus subtilis]	54	26	816
220	1	1322	663	100	gi 1333761	100	myosin II heavy chain [Neogleria fowleri]	54	22	660
220	13	12655	13059	100	pir S00485 S004	100	gene 11-1 protein precursor - Plasmodium falciparum (fragments)	54	35	405
221	3	2030	3709	100	gi 1303813	100	Yqew [Bacillus subtilis]	54	34	1680
272	7	5055	4219	100	gi 62964	100	arylamine N-acetyltransferase (AA 1-290) [Gallus gallus] ir S06652 MYCHV3	54	33	837
316	7	4141	4701	100	gi 682769	100	arylamine N-acetyltransferase (EC 2.3.1.5) (clone NAT-3) - chicken	54	31	561
316	10	6994	8742	100	gi 413951	100	accE gene product [Escherichia coli]	54	28	1749
338	3	3377	2214	100	gi 450328	100	lpa-27d gene product [Bacillus subtilis]	54	28	1163
341	4	3201	3614	100	gi 171959	100	LORF F [unidentified]	54	25	414
							myosin-like protein [Saccharomyces cerevisiae]			

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
346	1	1820	912	gi 396400	similar to eukaryotic Na <sup>+</sup> /H <sup>+</sup> exchangers [Escherichia coli] sp P32703 YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOXR-ACS HYPERGENIC REGION (0549)	54	34	909
348	2	623	1351	gi 537109	ORF_5343a [Escherichia coli]	54	34	729
378	2	1007	1942	sp P02983 TCR_S	TETRACYCLINE RESISTANCE PROTEIN	54	31	936
408	6	4351	5301	gi 474190	lucA gene product [Escherichia coli]	54	29	951
444	9	7934	8854	gi 216267	ORF2 [Bacillus megaterium]	54	32	921
463	2	2717	2229	gi 304160	product unknown [Bacillus subtilis]	54	50	489
502	2	1696	1133	gi 1205015	hypothetical protein (SP10120) [Haemophilus influenzae]	54	38	564
505	6	6262	5357	gi 1500558	2-hydroxyhepta-2,4-diene-1,7-diole isomerase (Methanococcus jannaschii)	54	41	906
550	1	2736	1522	gi 40100	rodC (tag3) polypeptide (AA 1-746) [Bacillus subtilis] tr S06049 S06049 rodC protein [Bacillus subtilis] p P33685 TAGP_BACSU TECHOIC ACID BIOSYNTHESIS PROTEIN F.	54	35	1215
551	5	1305	4279	gi 950197	unknown [Corynebacterium glutamicum]	54	34	975
558	2	1356	958	gi 485090	No definition line found [Caenorhabditis elegans]	54	32	399
580	1	91	936	gi 331906	fused envelope glycoprotein precursor [Friend spleen focus-forming virus]	54	45	846
603	3	554	757	gi 1323423	ORF YGR234w [Saccharomyces cerevisiae]	54	36	204
617	1	25	249	gi 219959	ornithine transcarbamylase [Homo sapiens]	54	40	225
622	3	1097	1480	gi 1303873	YggZ [Bacillus subtilis]	54	25	384
623	1	3	404	gi 1063250	low homology to P20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillius [Bacillus subtilis]	54	45	402
689	1	1547	1011	gi 552446	NADH dehydrogenase subunit 4 [Apis mellifera ligustical] pir S22968 S22968 NADH dehydrogenase chain 4 - honeybee mitochondrion (S024)	54	30	537
725	2	686	1441	gi 987096	sensory protein kinase [Streptomyces hygroscopicus]	54	26	756
956	1	1	249	pir S30782 S307	integrin homolog - yeast [Saccharomyces cerevisiae]	54	24	249
978	2	1137	859	gi 1301994	ORF YNL091w [Saccharomyces cerevisiae]	54	33	279
1314	1	3	281	gi 1001108	hypothetical protein [Synchocystis sp.]	54	33	279
2450	1	1	228	gi 1045057	ch-TOG [Homo sapiens]	54	32	228
2934	1	1	387	gi 580870	ipa-37d qoxA gene product [Bacillus subtilis]	54	36	387
2970	1	499	251	sp P3734P VECE	HYPOTHETICAL PROTEIN IN ASP5 5' REGION (FRAGMENT)	54	42	249

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3002	1	1	309	gi 44027	Taa protein [Lactococcus lactis]	54	33	309
3561	1	9	464	gi 151259	HMG-CoA reductase (EC 1.1.1.89) [Pseudomonas mvaloniis] pir[A44756]A44756 hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) Pseudomonas sp.	54	35	456
3572	1	72	401	gi 450686	hcdM gene of Ecoprr1 gene product [Escherichia coli] pir[S38437]S38437 hcdM protein - Escherichia coli pir[S09629]S09629 hypothetical protein A - Escherichia coli (SUB 40-520)	54	36	330
3829	1	798	400	gi 1322245	mevalonate pyrophosphate decarboxylase [Rattus norvegicus]	54	29	399
3909	1	1	273	gi 29865	CENP-E (Homo sapiens)	54	30	273
3921	1	3	209	pir[S24325]S243	glucan 1,4-beta-glucosidase (EC 3.2.1.74) - Pseudomonas fluorescens subsp. cellulosa	54	34	207
4438	1	566	285	gi 1196657	unknown protein [Mycoplasma pneumoniae]	54	30	282
4459	1	3	272	gi 1046081	hypothetical protein (GB:D26185_10) [Mycoplasma genitalium]	54	38	270
4564	1	3	221	gi 216267	ORF2 [Bacillus megaterium]	54	38	219
23	12	12538	10685	gi 474192	lucC gene product [Escherichia coli]	53	35	1854
23	14	14841	13579	gi 42029	ORF1 gene product [Escherichia coli]	53	32	1263
24	3	4440	3940	gi 1369947	c2 gene product [Bacteriophage B1]	53	36	501
26	4	3818	4618	gi 1486247	unknown [Bacillus subtilis]	53	37	801
34	6	2856	3998	gi 405880	yell [Escherichia coli]	53	40	1143
38	10	9380	7806	gi 1399954	thyroid sodium/iodide symporter NIS [Rattus norvegicus]	53	29	1575
56	10	12324	12100	pir[A54592]A545	110k actin filament associated protein - chicken	53	32	225
57	6	5047	4583	pir[A00341]DE2P	alcohol dehydrogenase (EC 1.1.1.1) - fission yeast [Schizosaccharomyces pombe]	53	39	465
57	12	10515	8932	gi 480429	putative transcriptional regulator [Bacillus stearothermophilus]	53	30	1584
67	12	9496	10218	gi 1511555	quinolone resistance nraA protein protein [Methanococcus jannaschii]	53	31	723
69	3	3125	2382	gi 1687017	arabinogalactan-protein, AGP [Nicotiana glauca, cell-suspension culture filtrate, Peptide, 461 aa]	53	30	744
79	1	3	1031	gi 1523802	glucanase [Anabaena variabilis]	53	32	1029
80	1	673	338	gi 452428	ATPase 3 [Plasmodium falciparum]	53	36	336
88	4	1910	2524	gi 137034	ORF_0488 [Escherichia coli]	53	25	615
88	5	2467	3282	gi 537034	ORF_0488 [Escherichia coli]	53	29	816

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match	Gene name	% sim	% ident	length (nt)
92	8	5870	5505	gi 199598	amphotropic murine retrovirus receptor [Rattus norvegicus]	53	33	366
94	5	4417	3239	gi 173038	tropomyosin (TPM1) [Saccharomyces cerevisiae]	53	25	1179
99	5	4207	5433	sp P28246 BCR_E	BICYCLOMYCIN RESISTANCE PROTEIN (SULFONAMIDE RESISTANCE PROTEIN)	53	30	1227
120	3	1639	2262	gi 576655	ORF1 [Vibrio anguillarum]	53	35	624
120	11	7257	8897	gi 1524397	glycine betaine transporter Opud [Bacillus subtilis]	53	33	1641
127	6	6893	5685	gi 1256630	putative [Bacillus subtilis]	53	32	1209
147	2	255	557	gi 581648	epib gene product [Staphylococcus epidermidis]	53	36	303
158	4	4705	4256	gi 151004	mucoid regulatory protein Algr [Pseudomonas aeruginosa] pIr A32802 A32802 regulatory protein algr - Pseudomonas aeruginosa sp P26275 ALCR_PSEAB POSITIVE ALGINATE BIOSYNTHESIS REGULATORY PROTEIN	53	32	450
171	7	5717	5421	gi 1510669	hypothetical protein (GP:D64044_18) [Methanococcus jannaschii]	53	34	297
191	9	13087	11483	gi 298085	acetate decarboxylase [Clostridium acetobutylicum] pIr B49346 B49346 butyrate--acetate CoA-transferase (EC 2.7.2.8) small chain - Clostridium acetobutylicum sp P33752 CTFA_CLOAB BUTYRATE-ACETOACETATE CO-TRANSFERASE SUBUNIT (EC 2.8.3.9) (COAT A)	53	31	1605
203	5	3763	4326	gi 143456	lipoB protein ltrg start codon [Bacillus subtilis]	53	29	564
206	17	18204	18971	gi 304136	acetylglutamate kinase [Bacillus stearothermophilus] sp Q07905 ARGB_BACST ACETYLGLUTAMATE KINASE (EC 2.7.2.8) (NAG INASE) (ACK) [N-ACETYL-L-GLUTAMATE 5-PHOSPHOTRANSFERASE]	53	36	768
212	10	4021	4221	gi 19878	protein kinase [Plasmodium falciparum]	53	28	201
231	2	1580	1350	gi 537506	paramyosin [Drosophila imitator]	53	34	231
272	6	2719	3249	pir A33141 A331	hypothetical protein (gtd 3' region) - Streptococcus mutans	53	34	531
308	3	927	2576	gi 606292	ORF_0696 [Escherichia coli]	53	33	1650
320	7	5645	5884	gi 160596	RNA polymerase III largest subunit [Plasmodium falciparum] sp P27625 PFC1_PLAFA DNA-DIRECTED RNA POLYMERASE III LARGEST SUBUNIT (EC 2.7.7.6)	53	33	240
327	1	218	901	gi 854601	unknown [Schizosaccharomyces pombe]	53	31	684
341	2	212	2500	gi 633732	ORF1 [Campylobacter jejuni]	53	31	2289
351	1	763	383	sp P31675 YABM_	HYPOTHETICAL 42.7 KD PROTEIN IN TUBA-LEUD INTERGENIC REGION (ORF104)	53	32	381
433	7	5087	4731	gi 1001961	MHC class II analog [Staphylococcus aureus]	53	30	357
454	2	1240	980	pir A60128 A603	40K cell wall protein precursor (sr 5' region) - Streptococcus mutans (strain OM5175, serotype f)	53	27	261

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Accession gene name	% sim	% ident	length (nt)
470	4	1123	1761	gi 516826	rat GCP360 (Rattus rattus)	53	30	639
483	1	432	217	gi 1480429	putative transcriptional regulator (Bacillus stearothermophilus)	53	33	216
544	1	516	1259	gi 46587	ORF 1 (AA 1 - 121) (1 to 2nd base in codon) (Staphylococcus aureus) irf1576515765 hypothetical protein 1 (h1b 5' region) - ephylococcus aureus (fragment)	53	38	744
558	10	3957	3754	gi 151140	res gene (Bacteriophage P1)	53	32	204
603	2	339	620	gi 507738	Hsp (Vibrio parahaemolyticus)	53	26	282
693	1	1669	941	gi 153123	toxic shock syndrome toxin-1 precursor (Staphylococcus aureus) p1r1A24606 KCSA1 toxic shock syndrome toxin-1 precursor - taphylococcus aureus	53	38	729
766	1	2	673	gi 487600	orfA2: orfA2 forms an operon with orfA1 (Listeria monocytogenes)	53	43	672
781	1	667	335	gi 1204551	pilin biogenesis protein (Haemophilus influenzae)	53	26	333
801	1	3	545	gi 1279400	SepA protein (Escherichia coli)	53	25	543
803	1	2	910	gi 695278	lipase-like enzyme (Alcaligenes eutrophus)	53	30	909
872	1	1177	590	gi 298032	EP (Streptococcus suis)	53	30	588
910	1	2	184	gi 1044936	unknown (Schistosoma haematodes)	53	29	183
943	1	794	399	gi 190508	similar to unidentified ORF near 47 minutes (Escherichia coli) ap P1416 YICK_ECOLI HYPOTHETICAL 43.5 KD PROTEIN IN SELC-NLPA NTERGENIC REGION.	53	30	396
988	1	1004	504	gi 142441	ORF 3; putative (Bacillus subtilis)	53	28	501
1064	1	3	434	gi 305080	myosin heavy chain (Entamoeba histolytica)	53	26	432
1366	1	3	452	gi 308852	transmembrane protein (Lactococcus lactis)	53	33	450
1758	1	792	397	gi 1001774	hypothetical protein (Synchocystis sp.)	53	30	396
1897	1	1	447	gi 1303949	YqIX (Bacillus subtilis)	53	27	447
2381	1	798	400	gi 1146243	22.4% identity with Escherichia coli DNA-damage inducible protein ...; putative (Bacillus subtilis)	53	37	399
3537	1	1	327	gi 450688	hadM gene of EcoPr1 gene product (Escherichia coli) p1r1S38437 S38437 hadM protein - Escherichia coli p1r1S09629 S09629 hypothetical protein A - Escherichia coli (sub 40-520)	53	35	327
3747	2	137	397	gi 1477486	transposase (Burkholderia cepacia)	53	53	261
11	5	3049	3441	gi 868224	no definition line found (Caenorhabditis elegans)	52	33	393

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
15	5	2205	2369	gi 215966	G41 protein (gag start codon) [Bacteriophage T4]	52	34	165
19	3	2429	3808	gi 1205379	UDP-muracil-pentapeptide synthetase [Haemophilus influenzae]	52	31	1380
24	1	6920	3462	gi 579124	predicted 85.4kd protein; 52kd observed [Mycobacteriophage L5] p1r[S10971]S10971 gene 26 protein - Mycobacterium phage L5 ap[Q05231]UG26 BPML5 MINOR TAIL PROTEIN GP26. (SUB 2-837)	52	32	3459
37	5	3015	3935	gi 1500543	P115 protein [Methanococcus jannaschii]	52	25	921
38	13	8795	9703	gi 46851	glucose kinase [Streptomyces coelicolor]	52	29	909
44	16	10817	11066	gi 42012	mosE gene product [Escherichia coli]	52	36	450
46	1	3	521	gi 1040957	NADH dehydrogenase subunit 6 [Anopheles tritarsus]	52	25	519
51	10	5531	6280	gi 388269	traC [Plasmid pAD1]	52	32	750
56	5	3968	2826	gi 181949	endothelial differentiation protein (edg-1) (Homo sapiens) p1r[A33100]A33100 C protein-coupled receptor edg-1 - human ap[P21453]EDG1 HUMAN PROBABLE G PROTEIN-COUPLED RECEPTOR EDG-1.	52	23	1143
57	5	4850	4173	gi 304153	sorbitol dehydrogenase [Bacillus subtilis]	52	27	678
62	5	3364	2870	gi 1072399	phas gene product [Rhizobium meliloti]	52	25	495
62	6	4445	3651	gi 46485	NADH dehydrogenase [Synechococcus PCC7942]	52	27	795
67	14	11355	12962	gi 1511365	glutamate synthase (NADPH), subunit alpha [Methanococcus jannaschii]	52	30	1608
67	21	116935	18158	gi 1204393	hypothetical protein (SP:p1122) [Haemophilus influenzae]	52	25	1224
70	4	2185	1997	gi 7227	cytoplasmic dynein heavy chain [Dictyostelium discoideum] r[A44357]A44357 dynein heavy chain, cytosolic - slime mold cytoostelium discoideum	52	36	189
96	10	10005	10664	gi 1408485	8650 gene product [Bacillus subtilis]	52	26	660
103	5	3986	3351	gi 1009368	respiratory nitrate reductase [Bacillus subtilis]	52	42	636
109	3	4102	3350	gi 699274	lmbE gene product [Mycobacterium leprae]	52	39	753
109	19	15732	17300	gi 3526981	amino acid permease YeeF like protein [Salmonella typhimurium]	52	30	1569
121	3	1412	981	gi 72931	unknown [Saccharomyces cerevisiae]	52	32	432
125	3	865	1680	gi 1296975	put gene product [Porphyromonas gingivalis]	52	38	816
130	2	659	1807	gi 1256634	25.8% identity over 120 aa with the Synanococcus sp. MpaV protein; putative [Bacillus subtilis]	52	36	1149
149	1	1164	593	gi 1225943	PBE3 terminase [Bacillus subtilis]	52	33	582
149	14	4687	4415	gi 1510368	M. jannaschii predicted coding region MJO272 [Methanococcus jannaschii]	52	35	273

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
167	1	216	1001	gi1146025	cell division protein [Escherichia coli]	52	43	786
188	1	120	1256	gi1474915	orf 337; translated orf similarity to SW: BCR_ECOLI bicyclomycin resistance protein of Escherichia coli [Coxiella burnetii] pir[54420]/st4207 hypothetical protein 337 - Coxiella burnetii (SUB -338)	52	26	1137
195	9	9161	8760	gi11028	mitochondrial outer membrane 72K protein (Neurospora crassa) r[A36682]A36682 72K mitochondrial outer membrane protein - rospora crassa	52	25	402
200	3	2065	2607	gi1142439	ATP-dependent nuclease [Bacillus subtilis]	52	35	543
203	4	2776	3684	gi1103698	BLTD [Bacillus subtilis]	52	25	909
227	8	5250	5651	gi1105080	myosin heavy chain (Entamoeba histolytica)	52	24	402
242	1	21	1424	gi11060877	BmrY [Escherichia coli]	52	32	1404
249	5	4526	4753	pir[C37222]C372	cytochrome P450 1A1, hepatic - dog (fragment)	52	23	228
255	1	2107	1055	gi1143290	penicillin-binding protein [Bacillus subtilis]	52	28	1053
276	7	3963	3664	gi11001610	hypothetical protein [Synchocystis sp.]	52	30	300
276	8	4456	4055	gi1416235	orf L3 [Mycoplasma capricolum]	52	26	402
289	2	1856	1449	gi1150900	CTP phosphohydrolase [Proteus vulgaris]	52	34	408
325	1	1	279	gi11204874	polypeptide deformylase (formylmethionine deformylase) [Haemophilus influenzae]	52	33	279
340	1	2017	1010	gi11215695	peptide transport system protein SapF homolog; SapF homolog [Mycoplasma pneumoniae]	52	33	1004
375	3	340	1878	gi1467446	similar to SpoVB [Bacillus subtilis]	52	28	1539
424	4	4104	3262	gi11478239	unknown [Mycobacterium tuberculosis]	52	34	843
430	1	3	575	pir[A42606]A426	orfA 3' to orf405 - Saccharopolyspora erythraea (fragment)	52	28	573
444	4	4728	3712	gi11408494	homologous to penicillin acylase [Bacillus subtilis]	52	31	1017
465	1	1802	903	gi1143331	alkaline phosphatase regulatory protein [Bacillus subtilis] pir[A27650]A27650 regulatory protein phoR - Bacillus subtilis sp[P23545]PHOR_BACSU ALKALINE PHOSPHATASE SYNTHESIS SENSOR PROTEIN MOR (BC 2.7.3.-1)	52	36	900
489	5	4705	4169	gi1755152	highly hydrophobic integral membrane protein [Bacillus subtilis] sp[P42953]TAGG_BACSU TEICHOIC ACID TRANSLOCATION PEPIDASE PROTEIN AGG.	52	32	537
495	1	1262	633	gi11204607	transcription activator [Haemophilus influenzae]	52	25	630
505	7	6004	5762	gi1142440	ATP-dependent nuclease [Bacillus subtilis]	52	28	243

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
517	2	1162	1614	gi 166162	Bacteriophage phi-11 int gene activator [Staphylococcus acteriophage phi 11]	52	35	453
543	2	444	1295	gi 1215693	putative orf; GT9_orf434 [Mycoplasma pneumoniae]	52	25	852
586	1	1	336	gi 581649	epiB gene product [Staphylococcus epidermidis]	52	36	336
773	1	848	426	gi 1279769	FDHC [Methanobacterium thermoformicicum]	52	30	423
1120	2	100	330	gi 142439	ATP-dependent nuclease [Bacillus subtilis]	52	35	231
1614	1	691	347	gi 289262	comS ORF3 [Bacillus subtilis]	52	28	345
2495	1	1	324	gi 216151	DNA polymerase (gene U; tgg start codon) [Bacteriophage SP02] gi 579197 SP02 DNA polymerase (aa 1-648) [Bacteriophage SP02] gi 421498 [DJBSP2 DNA-directed DNA polymerase (EC 2.7.7.7) - phage P02]	52	34	324
2931	1	566	285	gi 1256136	YbbG [Bacillus subtilis]	52	30	282
2943	1	577	320	gi 41713	hlaA ORF (AA 1-245) [Escherichia coli]	52	35	258
2993	1	588	295	gi 298032	EF [Streptococcus suis]	52	34	294
3667	1	612	307	gi 849025	hypothetical 64.7-kDa protein [Bacillus subtilis]	52	36	306
3944	1	478	260	gi 1218040	BAA [Bacillus licheniformis]	52	36	219
3954	2	613	347	gi 854064	I087 [Human herpesvirus 6]	52	50	267
3986	1	90	401	gi 1205919	[Na <sup>+</sup> and Cl <sup>-</sup> dependent gamma-aminobutyric acid transporter [Haemophilus influenzae]	52	33	312
4002	1	3	389	gi 40003	oxoglutarate dehydrogenase (NADP+) [Bacillus subtilis] pIP23129 [ODO1_BACSU 2-OXOGLOUTARATE DEHYDROGENASE E1 COMPONENT (EC 2.4.2) (ALPHA-KETOGLUTARATE DEHYDROGENASE)]	52	42	387
4020	1	1	249	gi 159388	ornithine decarboxylase [Leishmania donovani]	52	47	249
4098	1	438	220	gi 409795	No definition line found [Escherichia coli]	52	32	219
4248	1	3	212	gi 965077	Adp6 [Saccharomyces cerevisiae]	52	40	210
7	1	3	575	gi 895747	putative cel operon regulator [Bacillus subtilis]	51	28	573
21	4	2479	3276	gi 1510962	indole-3-glycerol phosphate synthase [Methanococcus jannaschii]	51	32	798
22	9	5301	5966	gi 1303933	YqiN [Bacillus subtilis]	51	25	666
43	3	1516	1283	gi 1519460	SrpI [Schizosaccharomyces pombe]	51	31	234
44	17	11042	11305	gi 42011	moaD gene product [Escherichia coli]	51	35	264
51	11	6453	6731	gi 495471	vacuolating toxin [Helicobacter pylori]	51	37	279

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
52	4	2537	2995	gi1256652	25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]	51	32	459
57	10	7331	6843	gi1508173	81% domain of PTS-dependent Cat transport and phosphorylation Escherichia coli	51	32	489
59	1	29	1111	gi1299163	alanine dehydrogenase (Bacillus subtilis)	51	33	1083
67	20	15791	16576	gi11510977	M. jannaschii predicted coding region M0938 [Methanococcus jannaschii]	51	24	786
69	2	1559	1218	gi1467359	unknown [Bacillus subtilis]	51	34	342
71	1	3	1196	gi1298032	EF [Streptococcus suis]	51	32	1194
78	2	349	176	gi1161242	proliferating cell nuclear antigen [Styela clava]	51	28	174
99	4	3357	4040	gi1642795	TFIID subunit TAF1155 [Homo sapiens]	51	25	684
109	1	2852	1428	gi1580920	rodD (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] p1r180648 S06048 probable rodD protein - Bacillus subtilis sp P13484 TAGE_BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPGA-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHNOIC ACID BIOSYNTHESIS ROTIN E)	51	27	1425
109	9	6007	6691	gi1204815	hypothetical protein (SP:P12662) [Haemophilus influenzae]	51	23	687
112	3	1066	2352	p1r1505330 S053	maltose-binding protein precursor - Enterobacter aerogenes	51	42	1287
112	13	14432	12855	gi1405857	yehU [Escherichia coli]	51	29	1578
114	9	9725	8967	gi1435098	orf1 [Mycoplasma capricolum]	51	30	759
115	1	1	912	gi11431110	ORF YH08JW [Saccharomyces cerevisiae]	51	28	912
127	10	9647	10477	gi1204314	H. influenzae predicted coding region HI0056 [Haemophilus influenzae]	51	37	831
152	9	6814	7356	gi1431929	Mun1 regulatory protein [Mycoplasma sp.]	51	38	543
154	2	575	1153	gi11237044	unknown [Mycobacterium tuberculosis]	51	36	579
154	7	6587	5634	gi1409286	BarU [Bacillus subtilis]	51	27	954
171	8	8943	6236	gi11205484	hypothetical protein (SP:P33918) [Haemophilus influenzae]	51	32	708
184	1	1	291	gi1466886	BarU [Bacillus subtilis]	51	33	291
212	5	1501	2139	p1r145605 A456	mature-parasite-infected erythrocyte surface antigen MESA - Plasmodium falciparum	51	23	639
228	2	707	1378	gi18204	nuclear protein [Drosophila melanogaster]	51	27	672
236	8	8137	7401	gi149272	Asparaginase [Bacillus licheniformis]	51	31	657
243	4	4637	3546	gi11511102	melvalonate kinase [Methanococcus jannaschii]	51	29	1092

TABLE 2



S aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
257	4	3540	3373	gi 204579	H. influenzae predicted coding region H10326 [Haemophilus influenzae]	51	22	168
258	3	2397	1609	gi 160299	glutamic acid-rich protein [Planodinium falcatum] pifAS6514	51	34	789
265	5	2419	3591	gi 580841	glutamic acid-rich protein precursor - Planodinium falcatum			
298	2	518	748	gi 1336162	glutamic acid-rich protein precursor - Planodinium falcatum	51	32	1173
316	9	5817	7049	gi 413953	SCP8 [Streptococcus agalactiae]	51	34	231
332	2	3775	2057	gi 1205012	ipa-29d gene product [Bacillus subtilis]	51	39	1233
364	4	3816	4991	gi 528991	mutS [Thermus aquaticus thermophilus]	51	26	1719
440	2	448	684	gi 2819	unknown [Bacillus subtilis]	51	32	1176
495	2	1353	1177	gi 297861	transferase [Glu10] (AA 1 - 687) [Kluyveromyces fragilis] r[S01407]XUVKG	51	32	237
495	3	2287	1718	gi 1513317	UDP-glucose 4-epimerase (EC 5.1.3.2) - yeast uveromyces marxianus var. lactis			
506	1	840	421	gi 455320	protease G [Erwinia chrysanthemi]	51	41	177
600	1	1474	983	gi 587532	serine rich protein [Entamoeba histolytica]	51	25	570
607	3	479	938	gi 1511524	icli protein [Bacteriophage P4]	51	33	420
686	2	127	600	gi 493017	orf. lon: 201. CAI: 0.16 [Saccharomyces cerevisiae] pif[S48818]S48818	51	30	492
726	1	33	230	gi 1353851	hypothetical protein - yeast [Saccharomyces cerevisiae]			
861	1	176	652	gi 410145	hypothetical protein (SP:P37002) [Methanococcus jannaschii]	51	40	456
869	1	782	393	gi 40100	endocarditis specific antigen [Enterococcus faecalis]	51	30	474
1003	1	642	322	gi 1279707	unknown [Prochlorococcus marinus]	51	45	198
1046	2	866	624	gi 510257	dehydroquinase dehydratase [Bacillus subtilis]	51	34	477
1467	1	702	352	gi 1511175	rodC [tag] polypeptide (AA 1-746) [Bacillus subtilis] it[S06049]S06049	51	23	390
2558	1	457	210	ep P10582 DROM_	rodC protein - Bacillus subtilis p P13485 TAGF_BACSU TECHNOIC ACID BIOSYNTHESIS PROTEIN P.			
3003	1	779	399	gi 809543	hypothetical phosphoglycerate mutase [Saccharomyces cerevisiae]	51	39	321
3604	1	1	399	p JC4210 JC42	glycosyltransferase [Escherichia coli]	51	29	243
3732	1	2	316	gi 145906	M. jannaschii predicted coding region MJ1177 [Methanococcus jannaschii]	51	32	351
					DNA POLYMERASE (EC 2.7.7.7) (S-1 DNA ORF 3).	51	26	228
					[ChrC protein [Erwinia chrysanthemi]]	51	27	381
					[3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35) - mouse	51	37	399
					acyl-CoA synthetase [Escherichia coli]	51	33	315

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3791	1	2	274	gi 1061351	semaphorin III family homolog (Homo sapiens)	51	37	273
3995	1	46	336	gi 216346	surfactin synthetase (Bacillus subtilis)	51	38	291
4193	1	612	307	gi 42749	ribosomal protein L12 (AA 1-179) (Escherichia coli) irjS04776 XXECPL peptide N-acetyltransferase rmlB EC 2.3.1.-) - chericchia coli	51	25	306
4539	1	367	185	gi 1408494	homologous to penicillin acylase (Bacillus subtilis)	51	40	183
4562	1	442	239	gi 1456280	coded for by C. elegans cDNA cm01e7; Similar to hydroxymethylglutaryl-CoA synthase (Caenorhabditis elegans)	51	35	204
1	4	3576	4859	gi 559160	GNATL score: null; cap site and late promoter motifs present putative (Autographa californica nuclear polyhedrosis virus)	50	44	1284
11	7	4044	5165	gi 1146207	putative (Bacillus subtilis)	50	35	1122
11	13	10509	9496	gi 1208451	hypothetical protein (Synecocystis sp.1)	50	39	1014
19	1	2034	1018	gi 413966	ipa-42d gene product (Bacillus subtilis)	50	29	1017
20	11	8586	8407	gi 1323159	ORF YGR103W (Saccharomyces cerevisiae)	50	28	180
24	5	5408	4824	gi 496280	structural protein (Bacteriophage Tuc2009)	50	29	585
34	4	1926	2759	gi 1303966	YqjO (Bacillus subtilis)	50	36	834
38	30	22865	23440	gi 1072179	Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) (Caenorhabditis elegans)	50	32	576
47	2	1705	2976	gi 153015	FemA protein (Staphylococcus aureus)	50	29	1272
56	13	15290	15841	gi 606096	ORF_f167; end overlaps end of o100 by 14 bases; start overlaps f174, ther starts possible (Escherichia coli)	50	30	562
57	1	2135	1077	gi 640922	xylitol dehydrogenase (unidentified hemiascomycete)	50	29	1059
58	2	628	1761	gi 143725	putative (Bacillus subtilis)	50	29	1134
88	6	4393	3884	gi 1072179	Similar to dihydroflavonol-4-reductase (maize, petunia, tomato) (Caenorhabditis elegans)	50	32	510
89	5	3700	3356	gi 1276658	ORF174 gene product (Porphyra purpurea)	50	25	345
141	1	3	239	gi 476024	carbamoyl phosphate synthetase II (Plasmodium falciparum)	50	33	237
151	1	186	626	gi 1403441	unknown (Mycobacterium tuberculosis)	50	35	441
166	7	11065	9623	gi 895747	putative cel operon regulator (Bacillus subtilis)	50	32	1443
201	6	5284	5096	gi 160129	circumsporozoite protein (Plasmodium reichenowi)	50	42	189
206	22	30784	29555	gi 1052754	LarP integral membrane protein (Lactococcus lactis)	50	24	1230

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
211	4	1523	1927	gi 410131	ORFX7 [Bacillus subtilis]	50	29	405
214	4	2411	3295	ep P37348 VECE_	HYPOTHETICAL PROTEIN IN ASP5 5'REGION (FRAGMENT)	50	37	885
228	7	5068	4406	gi 313580	envelope protein (Human immunodeficiency virus type 1) p1rIS5835[S35835 envelope protein - human immunodeficiency virus type 1 (fragment) (SUB 1-77)]	50	35	663
272	2	3048	1723	gi 1408485	B65G gene product [Bacillus subtilis]	50	22	1326
273	2	1616	984	gi 140186	phosphoglycerate mutase [Saccharomyces cerevisiae]	50	28	611
328	2	2507	1605	gi 148896	lipoprotein [Haemophilus influenzae]	50	26	903
332	4	5469	3802	gi 1526547	DNA polymerase family X [Thermus aquaticus]	50	27	1668
342	5	3473	3931	gi 456562	G-box binding factor [Dictyostellum discoideum]	50	35	459
352	1	1478	741	gi 288301	ORF2 gene product [Bacillus megaterium]	50	29	738
408	7	5299	5523	gi 111665	ORF2136 [Marchantia polymorpha]	50	27	225
420	3	650	1825	gi 1757842	UDP-sugar hydrolase [Escherichia coli]	50	30	1176
464	1	1	591	gi 487282	[Na <sup>+</sup> -ATPase subunit J [Enterococcus hirae]	50	29	591
472	2	1418	864	gi 551875	[BglR [Lactococcus lactis]	50	23	555
520	1	23	541	gi 567036	[CapE [Staphylococcus aureus]	50	27	519
529	1	6	410	gi 1256652	25% identity to the E.coli regulatory protein MprA; putative [Bacillus subtilis]	50	34	405
534	5	7726	6059	gi 1295671	selected as a weak suppressor of a mutant of the subunit AC40 of DNA dependent RNA polymerase I and III [Saccharomyces cerevisiae]	50	18	1668
647	1	2990	1497	gi 405568	TraI protein shares sequence similarity with a family of topoisomerases [Plasmid pSK41]	50	31	1494
664	3	1133	711	gi 410007	Leukocidin F component [Staphylococcus aureus, MRSA No. 4, Peptide, 23 aa]	50	32	423
678	1	1	627	gi 238032	EF [Streptococcus suis]	50	29	627
755	3	947	1171	gi 150572	cytochrome c1 precursor (EC 1.10.2.2) [Paracoccus denitrificans] gi 45465 cytochrome c1 (AA 1-450) [Paracoccus denitrificans] p1r[C29413][C29413 ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome c1 precursor - Paracoccus denitrificans sp P13627 CY1	50	37	225
827	1	1363	683	gi 142020	heterocyst differentiation protein [Anabaena sp.]	50	21	681
882	1	3	752	gi 408485	B65G gene product [Bacillus subtilis]	50	27	750
910	2	438	887	gi 104727	lysine-specific transport protein [Haemophilus influenzae]	50	25	450

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
933	1	524	760	gi1205451	cell division inhibitor [Haemophilus influenzae]	50	32	237
973	1	424	216	gi1886947	orf3 gene product [Saccharomyces cerevisiae]	50	40	189
1009	1	653	429	gi153727	M protein [group G streptococcus]	50	28	225
1027	1	511	257	gi1413934	ipa-10r gene product [Bacillus subtilis]	50	25	255
1153	2	556	376	gi1773676	nccA [Alcaligenes xyloxydans]	50	36	231
1222	1	798	400	gi1408485	B65G gene product [Bacillus subtilis]	50	21	399
1350	1	692	399	gi1289272	ferrichrome-binding protein [Bacillus subtilis]	50	32	294
2945	1	366	184	gi171704	hexaprenyl pyrophosphate synthetase (COQ1) [Saccharomyces erevisiae]	50	34	183
2968	2	1604	804	gi1397526	clumping factor [Staphylococcus aureus]	50	33	801
2998	2	657	394	gi1495696	F54E7.3 gene product [Caenorhabditis elegans]	50	40	264
3046	2	506	306	gi1513819 S138	acyl carrier protein - Anabaena variabilis (fragment)	50	32	201
3063	1	547	275	gi1474190	lucA gene product [Escherichia coli]	50	29	273
3174	1	3	146	gi1515900	alcohol dehydrogenase [Rhodobacter sphaeroides]	50	31	144
3792	1	625	314	gi1001423	hypothetical protein [Synecocystis sp.]	50	35	312
3800	1	2	262	gi1144733	NAD-dependent beta-hydroxybutyryl coenzyme A dehydrogenase Clostridium acetobutylicum	50	28	261
3946	1	173	188	gi1576765	cytochrome b [Myrmecia pilosula]	50	38	186
3984	1	578	291	lap137348 YECE_	HYPOTHETICAL PROTEIN IN ASP5 5'-REGION (FRAGMENT)	50	37	288
37	10	8250	7885	gi11204367	hypothetical protein (GB:014003_278) [Haemophilus influenzae]	49	30	366
46	16	13802	14848	gi1466860	lact: B1308_P1_34 [Mycobacterium leprae]	49	24	1047
59	5	2267	3601	gi1606304	ORF_0462 [Escherichia coli]	49	27	1335
112	18	17884	18615	gi1559502	MD4 protein (AA 1 - 409) [Caenorhabditis elegans]	49	25	732
138	9	6973	7902	gi1303953	esterase [Acinetobacter calcoaceticus]	49	29	930
217	6	4401	5138	gi1456254	fibrinectin/fibrinogen-binding protein [Streptococcus pyogenes]	49	31	738
220	12	11803	12657	gi1397526	clumping factor [Staphylococcus aureus]	49	31	855
228	4	1842	2492	pir1521692 S236	hypothetical protein 9 - Plasmodium falciparum	49	24	651
268	1	5016	2614	gi143047	ORF8 [Bacillus subtilis]	49	26	2403

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
271	2	1164	1373	gi 1001257	[hypothetical protein [Synecocystis sp.]	49	38	210
300	3	4340	3180	gi 1510796	[hypothetical protein [GP.X91006.2] [Methanococcus jannaschii]	49	26	1161
381	1	2281	1142	gi 396301	[saeches P50004: Bacterial regulatory proteins, araC family, ignature [Escherichia coli]	49	29	1140
466	1	3	947	gi 103863	[yqgp [Bacillus subtilis]	49	26	945
666	1	379	191	gi 633112	[ORF1 [Streptococcus sobrinus]	49	29	189
670	2	403	1014	gi 1122758	[unknown [Bacillus subtilis]	49	32	612
709	1	1433	795	gi 143830	[xpac [Bacillus subtilis]	49	29	639
831	1	943	473	gi 601786	[phosphonamoutase [Mycoplasma pirum]	49	29	471
1052	1	422	213	gi 1303799	[yqen [Bacillus subtilis]	49	21	210
1800	1	342	172	gi 216300	[peptidoglycan synthase enzyme [Bacillus subtilis] sp P37505[MURG_BACSU MURG PROTEIN UPD-N-ACETYLGLUCOSAMINE--N-ACETYLGLUCOSAMINE PENTAPEPTIDE]PYROPHOSPHORYL-UNDESCAPENOL N-ACETYLGLUCOSAMINE RANSEFERASE).	49	28	171
2430	1	2	376	sp P27434 YFCA..	[HYPOTHETICAL 36.2 KD PROTEIN IN NDK-OCPE INTERGENIC REGION.	49	26	375
3096	1	542	273	gi 516360	[surfactin synthetase [Bacillus subtilis]	49	25	270
32	4	3771	3100	gi 11217963	[hepatocyte nuclear factor 4 gamma [HNF4gamma] [Homo sapiens]	48	36	672
38	1	1	609	gi 205790	[H. influenzae predicted coding region H11555 [Haemophilus influenzae]	48	28	609
45	6	5021	6427	gi 1534267	[unknown [Mycobacterium tuberculosis]	48	20	1407
59	14	16346	31096	gi 1197336	[lmp3 protein [Mycoplasma hominis]	48	30	606
61	1	3	608	gi 1511555	[quinolone resistance norA protein [Methanococcus jannaschii]	48	29	336
61	3	3111	3646	gi 1303893	[yqhl [Bacillus subtilis]	48	25	318
114	1	98	415	gi 671708	[su(s) homolog; similar to Drosophila melanogaster suppressor of able [su(s)] protein, Swiss-Prot Accession Number P22293 [Drosophila virilis]	48	25	318
121	1	1131	610	gi 1314584	[unknown [Sphingomonas 588]	48	29	522
136	1	2014	1280	gi 1205968	[H. influenzae predicted coding region H11738 [Haemophilus influenzae]	48	23	735
171	10	8220	9557	gi 1208454	[hypothetical protein [Synecocystis sp.]	48	34	1338
175	1	3625	1814	gi 396400	[similar to eukaryotic Na+/H+ exchangers [Escherichia coli] sp P32703 YJCE_ECOLI HYPOTHETICAL 60.5 KD PROTEIN IN SOAR-ACS INTERGENIC REGION 10549).	48	29	1812
194	1	2	385	gi 1510493	[H. jannaschii predicted coding region M30419 [Methanococcus jannaschii]	48	25	384

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Unit ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
197	1	901	452	gi 1045714	spermidine/putrescine transport ATP-binding protein [Mycoplasma genitalium]	48	25	450
203	1	1	396	gi 940288	protein localized in the nucleoli of pea nuclei; ORF; putative P1aum [satium]	48	29	396
204	1	1363	698	gi 529202	No definition line found [Caenorhabditis elegans]	48	25	666
206	20	34815	27760	gi 511490	gramicidin S synthetase 2 [Bacillus brevis]	48	27	7056
212	1	2	166	gi 295899	nucleolin [Xenopus laevis]	48	34	165
220	10	12652	11426	gi 44073	SecY protein [Lactococcus lactis]	48	23	1227
243	6	6450	5491	gi 1184118	mevalonate kinase [Methanobacterium thermoautotrophicum]	48	30	960
264	4	5434	3208	gi 1015903	ORF YJ815c [Saccharomyces cerevisiae]	48	26	2127
441	1	1532	768	gi 142863	replication initiation protein [Bacillus subtilis] pir a26580 a26580 replication initiation protein - Bacillus subtilis	48	23	765
446	5	3498	5298	gi 145836	putative [Escherichia coli]	48	24	1401
484	2	368	1110	gi 146551	transmembrane protein (kdp) [Escherichia coli]	48	18	723
542	3	1425	2000	int 528969 5289	N-carbamoylserine amidohydrolase (EC 3.5.1.59) - Arthrobacter sp.	48	27	576
566	1	3	1019	gi 153490	tetracycline C resistance and export protein [Streptomyces laevis]	48	24	1017
611	1	2	730	gi 1103507	unknown [Schistosoma mansoni]	48	38	729
624	1	1255	665	gi 144859	ORF B [Clostridium perfringens]	48	26	591
846	1	1014	508	gi 537506	paraoxonin [Piroplasma leishmaniae]	48	27	507
1020	1	66	950	gi 1499476	mequnium and colalt transport protein [Methanococcus jannaschii]	48	30	885
1227	1	1	174	gi 493730	lipoygenase [Pinus sativum]	48	35	174
1266	1	1	405	gi 682452	ORF_f211; alternate name y99A; orf5 of X14436 [Escherichia coli] gi 41425 ORF5 (AA 1-197) [Escherichia coli] (EUB 15-311)	48	24	405
2071	1	707	381	gi 1408486	H574A gene product [Bacillus subtilis]	48	25	327
2398	1	463	233	gi 1500401	reverse gyrase [Methanococcus jannaschii]	48	40	211
2425	1	476	246	pir H48563 H485	G1 protein - fowlpox virus (strain HP444) (fragment)	48	40	211
2432	1	446	225	gi 1353703	Trlo [Homo sapiens]	48	33	222
2453	1	794	399	gi 142850	division initiation protein [Bacillus subtilis]	48	29	396
2998	1	469	236	gi 577569	pepV [Lactobacillus delbrueckii]	48	31	234

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	match gene name	% sim	% ident	length (nt)
3042	1	14	280	gi 945219	mucin [Homo sapiens]	48	35	267
3686	1	1	405	gi 145836	putative [Escherichia coli]	48	25	405
4027	2	492	301	pir_S51171SS11	trans-activator protein - Equine infectious anemia virus	48	32	192
4	2	3641	2232	gi 1303989	YqkI [Bacillus subtilis]	47	24	1410
24	3	599	1084	gi 1540083	PC4-1 gene product [Bradydia hygida]	47	28	486
36	10	7524	6925	gi 1209223	estorase [Acinetobacter lwoffii]	47	26	600
43	2	196	1884	gi 1403455	unknown [Mycobacterium tuberculosis]	47	27	1689
44	22	16118	15108	gi 1511555	quinolone resistance norA protein protein [Methanococcus jannaschii]	47	31	1011
69	7	7141	6710	gi 438466	Possible operon with orig. Hydrophilic, no homologue in the atabase; putative [Bacillus subtilis]	47	29	432
81	4	5022	4279	gi 466882	ppa1; 91496_C2_189 [Mycobacterium leprae]	47	24	744
120	12	9135	8863	gi 927340	D9509.27p; CAI: 0.12 [Saccharomyces cerevisiae]	47	38	273
142	1	2022	1174	gi 486143	ORF YK1034w [Saccharomyces cerevisiae]	47	32	849
168	1	2178	1093	gi 117254	hypothetical EcaB protein [Bacillus subtilis]	47	29	1086
263	1	1884	943	gi 142822	D-alanine racemase cds [Bacillus subtilis]	47	34	942
279	1	1109	561	gi 516608	2 predicted membrane helices. Homology with B. subtilis man Orf3 Rowland et. al. unpublished Accession number W74183, approximately 1 minutes on updated Rudd map; putative [Escherichia coli] sp P37355 YF8B_EC0011 HYPOTHETICAL 26.7 KD PROTEIN IN MEND-MENB	47	31	549
345	2	2620	1676	gi 1204835	hippuricase [Haemophilus influenzae]	47	28	945
389	2	152	400	gi 456562	G-box binding factor [Dictyostelium discoideum]	47	32	249
391	1	1	831	gi 1420856	myo-inositol transporter [Schizosaccharomyces pombe]	47	19	831
404	3	2072	2773	gi 1255425	C3JG8.2 gene product [Caenorhabditis elegans]	47	17	702
529	5	2145	3107	gi 1303973	YqjV [Bacillus subtilis]	47	29	963
565	2	2321	1257	gi 142824	processing protease [Bacillus subtilis]	47	28	1065
654	1	962	483	gi 243353	ORF 5' of ECRP3 [herpesvirus saimiri HVS, host-squirrel monkey, eptide, 407 aa]	47	23	480
692	1	115	633	gi 150756	40 kDa protein [Plasmid pJW1]	47	25	519
765	1	1634	819	gi 1256621	26.7% of identity in 165 aa to a Thermophilic bacterium hypothetical protein 6; putative [Bacillus subtilis]	47	28	816

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
825	2	211	1023	gi 397526	clumping factor (Staphylococcus aureus)	47	32	813
914	1	1	615	gi 558073	polymorphic antigen (Plasmodium falciparum)	47	29	615
1076	1	1	753	gi 1147557	aspartate aminotransferase (Bacillus circulans)	47	33	753
1351	1	793	398	gi 755153	ATP-binding protein (Bacillus subtilis)	47	20	396
4192	1	3	293	gi 145836	putative [Escherichia coli]	47	24	291
5	6	4708	4361	gi 305080	myosin heavy chain (Entamoeba histolytica)	46	30	348
11	4	2777	3058	gi 603639	Yel040p (Saccharomyces cerevisiae)	46	28	282
46	11	10518	10300	gi 1246901	ATP-dependent DNA ligase (Candida albicans)	46	28	219
61	4	3941	7930	gi 298032	EF (Streptococcus suis)	46	35	3990
132	4	5028	4093	gi 1511057	hypothetical protein SP-P45869 (Methanococcus jannaschii)	46	25	916
170	4	4719	3652	pir 551910 5519	G4 protein - Sauroleishmania tarentolae	46	26	1068
191	7	9543	8284	gi 1041334	P5405.7 (Caenorhabditis elegans)	46	25	1260
253	1	1	396	gi 1204449	dihydrofolate acetyltransferase (Haemophilus influenzae)	46	35	396
264	3	437	973	gi 180189	cerebellar-degeneration-related antigen (CDR34) (Homo sapiens) [gi 182737]	46	29	517
					cerebellar degeneration-associated protein (Homo sapiens) [pir A29770 A29770 cerebellar degeneration-related protein - human]			
273	1	485	285	gi 607573	envelope glycoprotein C1V3 region (Human immunodeficiency virus type 1)	46	35	201
350	1	3	563	gi 517052	ORF_f286 [Escherichia coli]	46	35	561
384	1	2	862	gi 121884	(urea?) amidolyase (Haemophilus influenzae)	46	31	861
410	4	1876	2490	gi 1110518	proton antiporter efflux pump [Mycobacterium smegmatis]	46	24	615
432	1	2663	1455	gi 1197634	orf4; putative transporter; Method: conceptual translation supplied by author [Mycobacterium smegmatis]	46	27	1209
458	1	2419	1211	gi 15470	portal protein [Bacteriophage SPP1]	46	30	1209
517	5	2477	4192	gi 1523812	orf5 [Bacteriophage A2]	46	23	1716
540	3	1512	1285	gi 215635	pacA [Bacteriophage P1]	46	30	228
587	2	649	1242	gi 537148	ORF_f181 [Escherichia coli]	46	29	594
1218	1	747	391	gi 1205456	single-stranded-DNA-specific exonuclease [Haemophilus influenzae]	46	30	357

TABLE 2



S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3645	1	1	402	gi 450888	hadK gene of Ecoprr1 gene product [Escherichia coli] p1r[S38437]S38437 hadK protein - Escherichia coli p1r[S09629]S09629 hypothetical protein A - Escherichia coli (SUS 40-520)	46	33	402
4176	1	673	338	gi 951460	FIM-C.1 gene product [Xenopus laevis]	46	31	336
37	7	4813	5922	gi 606064	ORF_408 [Escherichia coli]	45	24	1110
38	16	11699	12004	gi 452192	protein tyrosine phosphatase (PTP-BAS, type 2) [Homo sapiens]	45	24	306
87	2	1748	2407	gi 1064813	homologous to ap:PHOR-BACSU [Bacillus subtilis]	45	23	660
103	12	14182	13385	gi 1001307	hypothetical protein [Synecocystis sp.]	45	22	798
112	14	14791	13811	gi 1204389	H. influenzae predicted coding region H10131 [Haemophilus influenzae]	45	23	981
145	4	4483	3461	gi 1220578	open reading frame [Mus musculus]	45	20	1023
170	6	6329	4965	gi 238657	AppC-cytochrome d oxidase, subunit I homolog [Escherichia coli, K12, eptide, 514 aa]	45	27	1365
206	2	5230	4346	gi 1222056	aminotransferase [Haemophilus influenzae]	45	27	885
228	1	60	716	gi 160299	glutamic acid-rich protein [Plasmodium falciparum] p1r[A54514]A54514 glutamic acid-rich protein precursor - Plasmodium falciparum	45	23	657
288	1	2	1015	gi 1255425	G33C8.2 gene product [Caenorhabditis elegans]	45	23	1014
313	3	4339	3128	gi 1581140	NADH dehydrogenase [Escherichia coli]	45	30	1212
332	1	916	459	gi 870966	P47A4.2 [Caenorhabditis elegans]	45	20	456
344	1	3	221	gi 171225	kinesin-related protein [Saccharomyces cerevisiae]	45	26	219
441	2	1501	1073	gi 142863	replication initiation protein [Bacillus subtilis] p1r[B26580]B26580 replication initiation protein - Bacillus subtilis	45	27	429
672	1	2	982	gi 1511334	M. jannaschii predicted coding region MJ1323 [Methanococcus jannaschii]	45	22	981
763	3	1345	851	gi 606180	ORF_310 [Escherichia coli]	45	24	495
886	3	379	846	gi 726426	similar to protein kinases and C. elegans proteins F37C12.8 and 37C12.5 [Caenorhabditis elegans]	45	30	468
948	1	3	473	gi 156400	myosin heavy chain (isozyme unc-54) [Caenorhabditis elegans] p1r[A32958]MKW myosin heavy chain B - Caenorhabditis elegans sp P02566 MYB_CABEL MYOSIN HEAVY CHAIN B (HNC B)	45	25	471
1158	1	2	376	gi 441155	transmission-blocking target antigen [Plasmodium falciparum]	45	35	375
2551	1	4	285	gi 1276705	ORF287 gene product [Porphyra purpurea]	45	28	282
3967	1	42	374	gi 976025	HraA [Escherichia coli]	45	28	333

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% Ident	length (nt)
52	7	6931	5846	gi1467378	unknown [Bacillus subtilis]	44	22	1086
138	8	6475	6849	gi173028	thioredoxin II [Saccharomyces cerevisiae]	44	28	375
221	5	7032	5617	gi153490	tetracycline C resistance and export protein [Streptomyces lauecensis]	44	21	1416
252	2	1331	1122	gi1204989	hypothetical protein [GB-U00022.9] [Haemophilus influenzae]	44	30	210
263	2	1265	2093	gi1136221	carboxypeptidase [Sulfolobus solfataricus]	44	26	1173
365	4	4963	3524	gi11296822	orf1 gene product [Lactobacillus helveticus]	44	31	1440
543	3	1315	1833	gi1063250	low homology to P20 protein of Bacillus licheniformis and bleomycin acetyltransferase of Streptomyces verticillus [Bacillus subtilis]	44	24	519
544	4	3942	4892	gi1951460	FIM-C.1 gene product [Xenopus laevis]	44	32	951
792	1	1224	613	gi1205680	high molecular weight neurofilament [Rattus norvegicus]	44	28	612
44	14	11303	11911	gi1511614	polyhydroxybutyrate synthetase protein A [Methanococcus jannaschii]	43	27	609
59	8	3665	5128	gi1153490	tetracycline C resistance and export protein [Streptomyces lauecensis]	43	21	1464
59	10	5535	7527	gi1153022	lipase [Staphylococcus epidermidis]	43	22	1992
99	1	1346	681	gi11419051	unknown [Mycobacterium tuberculosis]	43	21	666
310	8	9402	12134	gi1397526	clumping factor [Staphylococcus aureus]	43	21	2733
412	3	2782	2303	gi1460540	sporozoite surface protein 2 - Plasmodium yoelii [fragment]	43	29	480
519	3	2547	3122	gi1006530	SULFIDE DEHYDROGENASE (FLAVOCYTOCHROME C) FLAVOPROTEIN CHAIN PRECURSOR (EC 1.8.2.-) (FC) (FCSU)	43	23	576
4	13	12053	13321	gi1295671	selected as a weak suppressor of a mutant of the subunit AC10 of DNA dependent RNA polymerase I and II [Saccharomyces cerevisiae]	42	18	1269
94	2	1768	1091	gi1501027	ORF2 [Trypanosoma brucei]	42	31	678
127	4	5791	4550	gi142029	ORF1 gene product [Escherichia coli]	42	21	1242
297	3	1515	1036	gi1142790	ORF1, putative [Bacillus firmus]	42	25	480
344	6	4097	1525	gi140320	ORF 2 (AA 1-203) [Bacillus thuringiensis]	42	30	573
512	1	2167	1115	gi1405957	yeef [Escherichia coli]	42	23	1053
631	1	2434	1223	gi1580920	rodd (gtaA) polypeptide (AA 1-673) [Bacillus subtilis] pir[S06048]S06048 probable rodd protein - Bacillus subtilis ap131484]TAGE.BACSU PROBABLE POLY(GLYCEROL-PHOSPHATE) LPH-GLUCOSYLTRANSFERASE (EC 2.4.1.52) (TECHOIC ACID BIOSYNTHESIS ROETEIN E)	42	24	1212

TABLE 2

S. aureus - Putative coding regions of novel proteins similar to known proteins

Cun	ORF	Start	Stop	match	match	gene name	% sim	% ident	length
ID	ID	(nt)	(nt)	accession	accession				(nt)
605	3	2359	1739	gi11303784	YqeD [Bacillus subtilis]		42	19	621
4132	1	787	395	gi1022910	protein tyrosine phosphatase [Dictyostelium discoideum]		42	25	393
86	2	1375	884	gi1309506	spermidine/spermine N1-acetyltransferase [Mus saxicola]		41	30	492
					spermidine/spermine N1-acetyltransferase - spiny ouse [Mus saxicola]				
191	12	14797	14075	gi1124957	orf4 gene product [Methanococcus jannaschii]		41	22	723
212	6	2150	3127	gi115873	observed 35.2kd protein [Mycobacteriophage 15]		41	26	978
213	3	1263	2000	gi1633692	TrsA [Yersinia enterocolitica]		41	18	738
408	4	2625	3386	gi11197634	orf4; putative transporter; Method: conceptual translation supplied by author [Mycobacterium mageritensis]		41	24	762
542	1	3	1103	gi1457146	rhostry protein [Plasmodium yoelii]		41	21	1101
924	1	2	475	pirJH0148JH01	nucleolin - rat		41	30	474
1562	1	1	402	gi1552184	asparagine-rich antigen Pf35-2 [Plasmodium falciparum]		40	20	402
					asparagine-rich antigen Pf35-2 - Plasmodium falciparum (fragment)				
2395	1	518	261	pirJ42251J422	hypothetical protein 5 - fowipox virus		40	18	258
4077	1	3	305	gi1055055	coded for by C. elegans cDNA yk37g1.5; coded for by C. elegans cDNA yk39g.5; alternatively spliced form of F52C9.8b [Caenorhabditis elegans]		39	21	303
958	1	1003	503	gi1255425	C33G8.2 gene product [Caenorhabditis elegans]		37	25	501
59	12	8294	10636	gi1535260	STAMP antigen [Plasmodium reichenowi]		36	24	2143
63	5	3550	8079	gi1298032	EF [Streptococcus suis]		36	19	4530
544	3	2507	3601	gi1015903	ORF YJRI1c [Saccharomyces cerevisiae]		35	22	1095
63	4	1949	3574	gi1552195	circumsporozoite protein [Plasmodium falciparum] sp[P050691]CSP_PLAFL CIRCUMSPOROZOITE PROTEIN (CS) (FRAGMENT)		32	27	1626

TABLE 2

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4	1	1234	692
4	3	1712	2278
4	4	3703	3032
4	14	13073	12585
5	2	2539	1601
5	3	1532	1771
5	7	4741	4550
5	9	7939	6422
5	12	8711	8547
6	4	2359	1982
8	1	349	176
11	8	5144	5983
11	9	5988	6498
11	10	6472	6284
11	16	10934	11271
12	5	5352	4942
12	6	4595	4862
15	3	1895	1650
16	10	11263	11085
18	2	1093	917
20	9	9125	7764
20	10	8571	8230
20	12	9201	8803
20	13	12158	10470
23	1	674	339
23	6	6138	5485
23	8	6376	5942

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
23	9	7651	6881
23	15	12618	12830
24	4	4556	4185
24	6	5642	5241
25	2	1824	2402
31	2	505	849
31	3	1177	1524
31	4	2454	3005
32	2	765	1388
32	9	7952	8575
32	10	8591	8728
32	11	9738	9379
32	12	10797	10087
34	2	1315	1049
36	7	5226	5801
36	13	7575	7261
36	12	7424	7621
37	4	3158	2964
38	2	1585	980
38	11	6425	6868
38	20	16982	16371
38	26	20253	20804
38	27	20732	21264
39	1	1	627
40	1	805	404
43	1	796	428
44	4	2674	2326

TABLE 3

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
44	5	2484	3263
44	14	10587	10129
44	20	13724	13536
44	21	13596	13994
45	7	6575	6297
46	8	6365	6520
46	12	10449	10976
46	17	15032	15424
47	1	288	1079
48	9	7620	7778
50	1	1612	962
50	2	1621	1316
51	1	738	370
51	5	2520	2245
53	1	442	287
53	7	6705	6319
54	7	9014	8709
55	1	592	326
55	3	1052	786
56	1	1	261
56	3	1551	1228
56	4	1970	1560
56	17	19092	18712
57	4	3694	3521
57	8	5436	5022
58	9	8885	8553
59	3	1366	1509

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
59	6	3026	2802
59	7	3770	3570
59	9	4946	4563
59	11	7518	8378
59	13	10401	116403
62	2	2656	1521
62	11	5440	5757
63	1	1	336
67	1	900	1781
67	2	1774	2610
67	3	2591	3904
67	8	7110	6955
68	1	78	326
70	6	6761	5199
70	11	8935	8645
77	3	1590	1192
79	2	1509	1228
79	3	1411	1791
81	1	2	403
85	9	8300	8653
85	10	8969	8781
86	3	1426	1232
87	8	9187	9366
88	3	1620	1922
89	1	3	161
89	7	5042	4878
91	1	1098	550

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
91	3	3938	3141
92	2	449	928
92	3	1958	1467
92	9	5638	6024
94	1	661	332
94	3	2445	1813
94	4	2583	2197
96	11	10601	11050
99	6	4672	4523
99	7	5014	4784
100	8	7658	7287
102	7	4697	4368
103	3	2496	2035
104	1	2	694
106	2	699	1277
105	1	1235	693
105	3	3233	2655
105	1	3	221
106	3	1209	1355
107	1	1081	542
109	4	4025	3651
109	13	11625	11996
109	14	11981	12268
109	20	17401	17680
110	1	2	760
114	10	8764	9384
116	1	1	309



TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig No	ORF ID	Start (nt)	Stop (nt)
	116	3	6273
	116	8	11049
	116	9	10313
	120	5	3703
	120	6	4270
	120	11	9290
	121	2	417
	126	3	1090
	127	3	2648
	127	5	4084
	131	6	6773
	132	2	715
	134	1	2
	135	2	512
	135	3	1124
	138	1	3
	138	7	6008
	140	1	2040
	140	2	2019
	140	5	2387
	142	2	1360
	142	7	8830
	143	7	7290
	146	1	1227
	146	1	2
	146	3	502
	146	4	3673

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Config ID	ORF ID	Start (nt)	Stop (nt)
	146	5	2874
	147	1	1
	149	11	3956
	149	12	4036
	149	13	4507
	149	15	4807
	149	16	5495
	149	18	5739
	149	21	7416
	149	23	9216
	149	24	9681
	149	25	10679
	150	2	2303
	154	3	1795
	154	8	6586
	154	14	12704
	154	15	13531
	156	1	315
	157	3	1183
	158	2	1471
	159	3	452
	161	2	876
	161	6	4653
	161	7	4803
	161	8	4896
	161	11	5817
	163	2	1604

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
163	5	2796	2304
163	7	2952	2607
163	9	4905	5132
164	3	1338	1147
166	3	5213	4854
168	4	2500	2868
168	5	3595	4158
170	3	2517	2777
171	2	2277	1450
171	11	12576	11125
172	1	3	278
172	2	1940	1149
173	1	1289	708
173	5	7001	6114
174	2	593	1105
175	3	2552	2890
175	5	1820	3315
175	7	4342	4306
182	4	5477	4986
184	5	6043	5702
188	2	1210	1755
188	4	2647	2994
189	6	2614	3039
190	3	1998	2564
191	1	1	153
191	2	950	669
191	10	11786	13039

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
191	111	12902	12363
192	1	91	426
195	3	2106	1932
195	5	2899	2606
198	2	1016	1591
201	1	170	625
203	2	783	1466
206	6	8930	7815
206	12	13947	13616
206	21	28208	27960
212	2	170	817
212	3	796	1167
212	7	3128	1436
212	9	3749	4075
213	1	1	705
214	2	1076	570
214	6	4064	3738
214	9	6600	6995
214	10	7864	7469
217	1	1927	965
218	1	178	657
218	3	1776	2156
220	2	1851	1369
220	3	3251	2262
220	7	8275	7208
220	8	10244	8661
220	9	11796	10216

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
221	4	1095	2613
221	9	11428	10757
226	1	3	659
226	2	2196	1459
226	3	1476	1961
227	1	2	487
227	2	460	975
227	4	1855	2121
227	5	2052	2345
227	6	4760	3768
227	9	5591	6367
228	5	2503	2877
228	6	2846	3526
233	7	3544	3762
236	2	809	579
238	2	1975	1391
239	2	1417	905
241	5	4495	4314
242	2	1677	1363
243	1	127	576
244	1	1291	647
244	2	3035	1962
245	2	1614	1258
246	1	69	215
246	4	738	1733
249	3	3906	3712
250	1	494	249

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
254	1	1	136
256	2	956	1144
257	3	3700	3227
260	4	4906	4580
261	4	2196	2606
261	6	3214	3681
264	2	155	439
264	5	5252	4533
264	6	4739	5107
267	2	1323	931
268	4	5140	4700
272	1	862	446
272	3	1200	1439
272	9	4691	4909
272	10	6489	6035
276	4	1746	1901
278	1	224	553
278	5	3299	3448
278	7	4849	5127
285	2	551	736
288	3	1756	1950
288	5	2055	2276
289	1	2107	1055
290	2	2234	1912
291	2	332	622
291	5	1545	2051
295	3	1606	1349

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
295	4	2728	2141
295	5	2220	2762
297	2	788	465
298	1	2	205
300	2	2380	1928
301	7	2794	2624
304	1	3	194
306	1	109	654
306	5	4036	4257
307	1	674	339
307	8	3645	3995
308	1	1	654
308	2	1120	599
308	4	2643	2332
313	2	2314	1919
314	1	10	702
316	2	982	1341
316	6	2758	3165
317	1	2	1114
317	3	4570	3458
321	6	5645	5217
321	7	6319	6140
321	8	7450	6794
322	2	827	543
326	2	165	1112
326	3	1117	1467
328	1	936	469

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
328	5	3452	3276
329	1	3	719
329	2	781	1212
329	3	1471	1833
330	1	576	289
330	3	1447	1623
332	3	2353	2204
332	7	4971	5138
333	2	3295	3128
335	1	864	433
337	2	95	526
340	2	1658	1356
341	1	3	281
341	3	2476	3192
341	5	3618	3944
341	6	3929	4558
344	5	3197	2889
345	1	1532	788
346	2	221	592
350	3	1410	1598
352	2	2178	1765
352	3	7316	4596
352	7	7967	8404
352	8	8906	9247
352	9	10171	9854
359	1	1	546
362	1	3	656



TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
364	2	2158	1808
364	8	10974	10714
365	2	1612	1313
365	5	4680	4090
365	7	4980	6219
366	3	520	1719
367	3	906	1085
368	1	748	494
375	1	2	136
380	3	1351	1097
389	1	1	276
390	1	2	877
390	2	1373	1549
391	2	751	580
395	1	391	197
396	1	2132	1068
398	3	1348	1141
399	1	176	669
401	3	566	847
402	2	100	465
404	8	5561	5370
408	2	1507	2369
408	3	2875	2672
408	5	3524	4423
410	3	2111	1890
413	1	880	488
416	1	607	320

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
416	2	578	867
416	3	2195	3590
417	1	3	179
417	2	161	616
420	2	788	513
422	2	357	677
431	2	858	1407
432	2	446	1084
433	1	1	417
433	3	2311	2033
434	1	942	535
434	2	2089	1235
440	1	1	450
442	2	1269	3320
443	3	1873	1520
444	1	1	696
444	7	6761	6366
451	1	940	614
453	2	896	636
453	8	3833	4786
453	9	4718	4512
453	10	4937	4731
455	1	434	219
455	2	472	930
459	1	265	687
462	1	2	247
466	2	1494	907

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
467	1	654	149
468	1	2	250
469	1	1488	925
469	3	2386	1372
469	4	3464	3706
470	1	77	538
470	6	4098	3694
470	7	6330	5686
470	9	7351	8181
470	10	8175	9733
471	1	940	500
471	2	1562	1017
476	1	70	267
477	1	2	760
477	3	1764	2081
477	4	2066	2332
480	5	4016	4261
481	2	956	480
486	3	613	774
487	6	1795	2112
488	1	715	359
492	1	127	675
493	1	2	520
493	2	496	1242
502	3	1149	1571
504	1	690	346
505	5	4566	4150

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
511	2	1741	1232
512	2	583	747
515	1	609	812
517	4	2179	2511
520	4	2097	2360
520	6	3908	3669
527	1	1	498
528	1	637	335
529	2	1679	1104
530	7	5298	5534
536	1	308	156
538	1	1362	736
538	3	2203	2480
538	5	3531	3121
538	6	4348	3731
540	1	996	864
540	2	1495	1031
541	1	89	433
541	2	719	432
542	2	1048	1272
545	2	1012	734
551	1	2145	1129
555	2	892	704
558	3	1357	1154
558	4	1760	1458
558	5	2105	1821
558	6	2166	2020

TABLE 3

TABLE 3

5. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
558	7	2636	2322
558	8	3053	2802
558	9	3986	3453
560	1	475	921
565	3	1706	1485
571	1	308	156
571	3	994	1206
577	1	2	199
577	2	163	453
579	1	1	477
579	2	1784	1200
583	1	1988	996
585	1	946	539
587	1	22	573
588	2	1896	1372
588	3	1742	1554
590	1	47	334
592	2	1455	1141
593	1	2	775
593	2	817	1122
595	1	87	890
596	3	1593	1435
602	1	8	169
603	5	1071	1469
606	1	322	768
607	5	1444	1226
610	1	1029	541

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Cpntig ID	ORF ID	Start (nt)	Stop (nt)
612	1	3	500
616	1	991	650
617	2	736	491
622	1	36	347
625	4	2046	2549
627	1	67	210
628	1	901	452
631	3	4789	4004
634	1	1448	759
636	1	189	368
636	2	1929	1063
637	2	2323	1994
638	1	227	1081
639	1	518	361
639	2	1377	811
641	1	118	444
642	3	1615	1331
642	4	2260	1847
643	1	3	608
645	4	1534	1758
645	6	2025	2321
645	7	2940	2488
648	1	2	1045
660	1	77	601
660	2	576	872
661	1	1725	961
664	2	89	304

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
667	1	3	411
668	1	1	330
671	2	812	516
673	1	3	338
674	2	865	584
679	1	1	237
679	3	1589	1906
688	1	1236	835
688	2	1352	1077
694	1	3	143
696	2	818	432
706	1	1	224
709	3	1183	1449
711	1	3	908
715	1	3	167
716	1	2	637
721	1	133	570
722	1	763	303
723	1	1656	829
723	2	1498	1112
727	1	2	472
729	1	268	641
731	1	130	828
735	1	2	214
736	1	3	782
738	1	2	298
742	1	3	230

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
745	1	1148	780
748	2	282	464
749	1	685	344
751	1	901	452
755	1	97	522
755	2	520	918
758	2	663	400
764	2	1033	746
767	1	1	405
768	1	2	373
771	1	1058	534
778	1	1735	902
785	1	1790	1023
787	1	1260	631
791	1	3	224
799	1	15	260
804	1	304	711
805	1	3	680
808	1	219	842
810	1	2221	1112
810	2	1774	1442
812	1	38	979
817	1	714	358
818	2	487	1104
819	2	1529	1012
819	3	1748	1419
820	1	195	1064

TABLE 3



TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
818	1	506	255
829	1	48	800
830	1	578	291
832	1	594	298
835	1	320	796
840	3	491	709
845	1	912	457
850	2	303	449
853	1	715	359
860	1	2	256
864	1	18	410
864	2	383	715
864	6	1676	1828
870	1	1	588
873	1	906	454
875	1	584	294
877	1	1661	1020
878	1	981	544
879	1	1567	785
881	1	1	243
882	1	389	604
890	1	2	508
905	1	793	398
906	1	852	544
912	1	373	188
913	1	3	290
913	2	1092	547

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
915	1	6	161
915	2	169	402
921	1	126	386
927	1	1578	808
928	1	2	385
929	1	2	400
932	1	2	400
934	1	1	384
936	1	1052	528
937	1	2	616
945	1	220	645
945	2	649	1242
946	1	1702	950
949	1	1	270
951	1	3	362
955	1	3	143
960	1	723	400
961	1	1	162
965	1	690	346
966	1	1079	606
969	1	3	302
971	1	12	170
974	1	319	161
976	1	692	348
977	1	2	211
982	1	1926	982
984	1	589	296

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
987	1	3	467
993	1	1	525
994	1	920	549
1006	1	557	318
1014	1	624	313
1015	1	2	463
1016	1	288	145
1019	1	1205	660
1022	1	839	474
1024	1	595	299
1024	2	276	431
1030	1	673	338
1032	1	355	179
1040	1	794	379
1043	1	3	269
1044	2	115	399
1047	1	1	159
1051	1	704	354
1051	2	1233	733
1063	1	2	400
1069	1	2	148
1069	2	769	533
1075	1	707	399
1077	1	97	405
1081	1	58	438
1086	1	1	384
1087	2	246	431

TABLE 3

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1088	1	3	374
1096	1	474	238
1098	1	1015	509
1100	1	1020	511
1100	2	1520	1158
1101	1	703	353
1102	1	385	194
1107	1	2	580
1114	1	3	422
1115	1	2	268
1119	1	22	267
1129	1	40	342
1132	1	360	181
1133	1	609	376
1144	1	446	225
1147	1	558	280
1153	1	1	153
1154	1	3	818
1159	1	1	330
1161	1	341	186
1164	1	427	256
1171	1	19	240
1171	2	108	299
1183	1	2	379
1195	1	355	179
1196	1	1	189
1200	1	33	197

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Config ID	ORF ID	Start (nt)	Stop (nt)
1203	2	129	464
1222	2	105	401
1232	1	1	387
1240	1	2	175
1247	1	520	311
1271	1	812	221
1286	1	2	595
1295	1	1	165
1306	1	367	185
1314	2	158	631
1316	1	58	570
1359	1	384	191
1370	1	1	402
1371	1	1	345
1374	1	710	357
1378	1	2	400
1392	1	3	413
1411	1	202	432
1433	1	311	167
1450	1	2	256
1453	1	295	149
1471	1	721	398
1477	1	869	639
1502	1	794	399
1518	1	126	449
1534	1	283	143
1546	1	3	401

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1547	1	506	255
1583	1	3	350
1587	1	3	563
1602	2	170	679
1629	1	1	402
1665	1	468	215
1760	1	625	314
1762	1	3	200
1876	2	119	286
1895	1	2	379
1931	1	798	400
1976	2	715	383
2055	2	252	401
2056	1	331	167
2150	1	523	263
2157	1	794	399
2164	1	564	283
2175	1	218	400
2212	1	492	331
2338	1	732	367
2342	1	3	167
2352	1	330	166
2352	2	622	398
2355	1	47	352
2356	1	679	341
2359	1	301	152
2421	1	296	150

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3046	1	367	185
3049	1	553	278
3050	1	3	314
3052	1	504	353
3065	1	2	157
3070	1	357	190
3075	1	440	222
3080	1	1	285
3092	1	320	162
3093	1	411	250
3100	1	52	237
3103	1	47	258
3118	1	344	174
3123	1	2	145
3127	1	1	147
3138	1	336	169
3142	1	388	203
3144	1	664	386
3151	1	337	170
3155	2	202	384
3168	1	12	176
3205	1	288	145
3282	1	1	150
3303	2	239	400
3371	2	211	399
3558	1	2	148
3558	2	36	401

TABLE 3

S. aureus - putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3046	1	367	185
3049	1	553	278
3050	1	3	314
3052	1	504	253
3065	1	2	157
3070	1	357	190
3075	1	460	222
3080	1	1	285
3092	1	320	162
3093	1	411	250
3100	1	52	237
3103	1	47	298
3116	1	344	174
3123	1	2	145
3127	1	1	147
3138	1	336	169
3142	1	388	203
3144	1	664	386
3151	1	337	170
3155	2	202	384
3168	1	12	176
3205	1	288	145
3282	1	1	150
3303	2	239	400
3371	2	211	399
3558	1	2	148
3558	2	36	401



TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3568	1	751	377
3595	1	757	380
3618	1	2	238
3618	2	130	402
3622	1	86	358
3622	2	664	398
3642	1	876	439
3649	1	781	398
3651	1	625	314
3664	1	467	637
3674	1	55	402
3677	1	619	311
3704	1	1	402
3726	1	535	269
3765	1	510	256
3779	1	554	357
3794	1	266	135
3794	2	667	377
3796	2	638	375
3801	1	474	262
3806	1	453	298
3807	1	42	389
3815	1	798	400
3827	1	3	320
3842	1	781	392
3853	1	671	399
3855	1	1	324

TABLE 3

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
3857	1	2	235
3861	1	590	297
3865	1	695	399
3897	1	3	173
3897	2	143	400
3898	2	225	401
3921	2	103	342
3927	1	70	375
3930	1	76	234
3946	2	651	382
3951	2	105	377
3965	1	646	344
3973	1	795	400
3981	1	3	311
3998	1	3	356
4001	1	481	296
4003	1	90	335
4018	1	2	259
4018	2	186	401
4021	1	1	345
4043	1	3	344
4054	1	3	344
4066	1	1	150
4070	1	1	324
4072	2	187	390
4073	1	1	285
4077	2	127	372

TABLE 3

*S. aureus* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4083	1	3	359
4090	1	27	368
4101	1	103	297
4105	1	1	306
4107	1	570	286
4119	1	629	319
4121	1	740	372
4123	1	3	230
4127	1	3	341
4128	1	2	331
4130	1	768	415
4146	1	97	381
4157	1	3	205
4186	1	505	254
4224	1	510	256
4239	1	1	348
4242	1	709	356
4252	1	589	296
4253	1	1	174
4256	1	568	323
4258	2	498	334
4267	1	284	144
4271	1	2	304
4287	1	303	163
4289	1	471	319
4302	1	153	305
4304	1	1	186

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4304	2	96	314
4306	1	2	151
4318	1	576	289
4322	1	5	148
4331	1	439	221
4331	2	528	364
4338	1	728	399
4346	1	671	277
4367	2	117	311
4373	1	2	288
4381	1	574	326
4384	1	614	309
4397	1	9	311
4402	1	1	249
4403	1	606	328
4406	1	3	317
4411	1	2	280
4411	2	697	398
4412	1	2	364
4418	1	3	230
4424	1	601	358
4443	1	427	215
4471	1	643	323
4478	1	540	271
4482	1	50	289
4489	1	601	302
4491	1	12	206

TABLE 3

TABLE 3

5

10

15

20

25

30

35

40

45

50

55

S. aureus - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
4495	1	3	179
4496	1	500	252
4500	1	130	306
4511	1	493	248
4518	1	1	246
4526	1	480	241
4527	1	2	163
4532	1	3	239
4542	1	11	175
4567	1	36	200
4573	1	1	231
4578	1	642	322
4619	1	1	180
4620	1	349	176
4662	1	1	246
4669	1	2	157
4680	1	28	183
4690	1	344	174

Table 4

ORF	SEQ ID NO	BLAST HOMOLOG	Antigenic Regions			
			Region 1	Region 2	Region 3	Region 4
168_6	5192	lipoprotein	36-45	84-103	152-161	176-185
238_1	5193	chrA	21-39	48-58	84-95	232-249
51_2	5194	OppB gene product (B. sub)	20-36	70-79	100-112	121-131
278_3	5195	lipoprotein 1	20-29	59-73	85-97	162-171
276_2	5196	lipoprotein	21-33	65-74	177-186	211-220
45_4	5197	ProX	28-37	59-69	85-100	120-129
315_8	5198	hypothetical protein	45-54	88-97	182-192	243-253
154_15	5199	unknown	31-40	48-58	79-88	95-104
228_3	5200	unknown	25-38	40-52	64-74	80-89
228_6	5201	unknown	29-41	89-101	128-143	173-184
50_1	5202	unknown	21-33	52-61	168-182	197-206
112_7	5203	iron-binding periplasmic	21-31	58-67	92-101	111-120
442_1	5204	unknown	30-39	91-100	122-137	182-192
66_2	5205	unknown	50-59	104-116	127-136	167-182
304_2	5206	Q-binding periplasmic	19-28	48-57	75-84	103-116
44_1	5207	hypothetical protein	27-36	86-95	129-138	192-201
161_4	5208	SphX	27-44	149-161	166-175	201-210
46_5	5209	cmpC (permease)	21-33	61-70	83-92	100-109
942_1	5210	traH [Plasmid pSK41]	83-92	109-118	127-142	
5_4	5211	ORF (S. aureus)	12-22	87-96	111-120	151-160
20_4	5212	peptidoglycan hydrolase (S.	24-34	129-138	141-150	161-171
328_2	5213	lipoprotein (H. flu)	81-90	123-133	290-299	
520_2	5214	fibronectin binding protein	44-54	63-79	81-90	95-110
771_1	5215	emm1 gene product (S. pyi	30-39	65-82	96-106	112-121
999_1	5216	predicted trithorax prot. (D	7-16	120-129	157-166	
853_1	5217	ORF2136 (Marchantia polyr	43-52	88-97	102-111	
287_1	5218	psaA homolog	13-22	28-44	72-82	114-124
288_2	5219	cell wall enzyme	14-23	89-98		
596_2	5220	penicillin binding protein 2b	40-49	59-68	76-87	106-115
217_5	5221	fibronectin/fibrinogen bindi	28-37	40-49	62-71	93-111
217_6	5222	fibronectin/fibrinogen bp	10-19	31-40	54-62	73-92
528_3	5223	myosin cross reactive prote	4-13	29-47	60-73	90-99
171_11	5224	EF	20-31	91-110		
63_4	5225	penicillin binding protein 2b	12-21	59-68	95-104	
353_2	5226		46-55	62-71		
743_1	5227	29 kDa protein in fimA regi	23-32	68-79	94-103	175-184
342_4	5228	Twitching motility	10-19	48-60	83-92	111-121
69_3	5229	arabinogalactan protein	97-106	132-141	158-167	180-189
70_6	5230	nodulin	36-45	48-57	137-160	179-188
129_2	5231	glycerol diester phosphodie	8-17	41-50	55-74	97-106
58_5	5232	PBP (S. aureus)	26-35	70-79	117-126	152-161
188_3	5233	MHC class II analog (S. aure	72-81	94-103	115-124	136-145
236_6	5234	histidine kinase domain (Dic	24-33	52-67	81-94	106-121
310_8	5235	clumping factor (S. aureus)	59-71	77-86	93-102	118-127
601_1	5236	novel antigen/ORF2 (S. aur	45-54	91-104	108-117	186-195
544_3	5237	ORF YJR151c (S. cerevisae)	76-90	101-111	131-140	154-164
662_1	5238	MHC class II analog (S. aure	22-32	71-80	89-98	114-122
87_7	5239	S' nucleotidase precursor ('	29-45	62-71	105-114	125-137
120_1	5240	B65G gene product (B. sub	102-111			

Table 4

ORF	Antigenic Regions		(cont)			
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
168_6	244-272	303-315				
238_1	260-269	291-301	308-317			
51_2	140-152	188-208	211-220	256-266	273-283	
278_3	198-209					
276_2	255-268					
45_4	177-199	221-230	234-243	268-279	284-293	304-313
316_8						
154_15	148-157	177-187	202-211			
228_3	101-119	139-154	166-181			
228_6						
50_1						
112_7	136-149	197-211	218-229	253-273		
442_1	199-210	247-257	264-277	287-309		
66_2						
304_2	178-187	250-259				
44_1						
161_4						
46_5	131-141	162-176	206-215	243-252	264-273	285-294
942_1						
5_4	189-205	230-239	246-264	301-318	340-354	378-387
20_4	202-212	217-234	260-275	314-336	366-373	380-391
328_2						
520_2						
771_1	145-154					
999_1						
853_1						
287_1	154-164					
288_2						
596_2	121-130					
217_5	244-253	259-268	288-297	302-311		
217_6	144-158	174-183	188-197	207-216	226-242	
528_3						
171_11						
63_4						
353_2						
743_1	197-207					
342_4						
69_3	195-211					
70_6	206-215	263-272	291-301	331-340	358-371	390-414
129_2	117-127	141-157	168-183	202-211	222-231	261-270
58_5	184-203	260-269	275-299	330-344	372-381	424-433
188_3						
236_6	138-147	163-172	187-198	244-261	268-278	308-317
310_8	131-140	144-153	177-186	190-199	204-213	216-227
601_1	208-218					
544_3	170-179	184-193	224-235	274-287	327-336	352-361
662_1						
87_7						
120_1						

Table 4

5	ORF	Antigenic		Regions		(cont)	
		Region 11	Region 12	Region 13	Region 14	Region 15	Region 16
	168_6						
	238_1						
	51_2						
10	278_3						
	276_2						
	45_4						
	316_8						
	154_15						
15	228_3						
	228_6						
	50_1						
	112_7						
	442_1						
20	66_2						
	304_2						
	44_1						
	161_4						
	46_5	306-315					
25	942_1						
	5_4	393-407	416-426	456-465			
	20_4	396-405	410-419	461-481			
	328_2						
	520_2						
30	771_1						
	999_1						
	853_1						
	287_1						
	288_2						
35	596_2						
	217_5						
	217_6						
	528_3						
40	171_11						
	63_4						
	353_2						
	743_1						
	342_4						
	69_3						
45	70_6	453-471	506-515				
	129_2	296-315					
	58_5						
	188_3						
50	236_6	358-377	410-423	428-439	442-457	467-476	480-493
	310_8	238-251	256-275	281-290	296-310	314-333	338-347
	601_1						
	544_3						
	662_1						
55	87_7						
	120_1						



EP 0 786 519 A2

Table 4

5	ORF	Antigenic Regions (cont)				
		Region 17	Region 18	Region 19	Region 20	Region 21
	168_6					
	238_1					
	51_2					
10	278_3					
	276_2					
	45_4					
	316_8					
	154_15					
15	228_3					
	228_6					
	50_1					
	112_7					
	442_1					
20	66_2					
	304_2					
	44_1					
	161_4					
	46_5					
25	942_1					
	5_4					
	20_4					
	328_2					
	520_2					
30	771_1					
	999_1					
	853_1					
	287_1					
	288_2					
35	596_2					
	217_5					
	217_6					
	528_3					
	171_11					
40	63_4					
	353_2					
	743_1					
	342_4					
	69_3					
45	70_6					
	129_2					
	58_5					
	188_3					
	236_6					
50	310_8	357-366	370-379	429-438	443-452	478-487
	601_1					551-560
	544_3					
	662_1					
55	87_7					
	120_1					

Table 4

	ORF	Antigenic Regions (cont)					
		Region 23	Region 24	Region 25	Region 26	Region 27	Region 28
5	168_6						
	238_1						
	51_2						
10	278_3						
	276_2						
	45_4						
	316_8						
	154_15						
15	228_3						
	228_6						
	50_1						
	112_7						
	442_1						
20	66_2						
	304_2						
	44_1						
	161_4						
	46_5						
25	942_1						
	5_1						
	20_4						
	328_2						
	520_2						
30	771_1						
	999_1						
	853_1						
	287_1						
	288_2						
35	596_2						
	217_5						
	217_6						
	528_3						
	171_11						
40	63_4						
	353_2						
	743_1						
	342_4						
	69_3						
45	70_6						
	129_2						
	58_5						
	188_3						
	236_6						
50	310_8	622-632	670-685	708-718	823-836	858-867	877-886
	601_1						
	544_3						
	662_1						
	87_7						
55	120_1						

Table 4

5

10

15

20

25

30

35

40

45

50

55

ORF	Antigenic Region 29	Regions Region 30	(cont)
168_6			
238_1			
51_2			
278_3			
276_2			
45_4			
316_8			
154_15			
228_3			
228_6			
50_1			
112_7			
442_1			
66_2			
304_2			
44_1			
161_4			
46_5			
942_1			
5_4			
20_4			
328_2			
520_2			
771_1			
999_1			
853_1			
287_1			
288_2			
596_2			
217_5			
217_6			
528_3			
171_11			
63_4			
353_2			
743_1			
342_4			
69_3			
70_6			
129_2			
58_5			
188_3			
236_6			
310_8			
601_1			
544_3			
662_1			
87_7			
120_1			

Table 4

ORF	BLAST		Antigenic Regions			
	HOMOLOG		Region 1	Region 2	Region 3	Region 4
46_1	5241	aldehyde dehydrogenase	8-17	36-52	83-96	112-121
63_4	5242	glycerol ester hydrolase (P.	9-26	57-73	93-107	123-133
174_6	5243	ketopantoate hydroxymeth	71-80	203-212	242-254	265-274
206_16	5244	ornithine acetyltransferase	1-10	34-43	54-63	194-210
267_1	5245	NaH-antiporter protein (E. f	120-129	332-347	398-408	
322_1	5246	acriflavin resistance protein	58-75	153-164	203-231	264-284
415_2	5247	transport ATP-binding prote	108-126	218-227	298-308	315-334
214_3	5248	2-nitropropane dioxygenase	123-136	216-233	283-292	297-306
587_3	5249	clumping factor	5-14	43-54	59-68	76-95
685_1	5250	signal peptidase	59-68	72-81	86-95	99-108
54_3	5251	fibronectin binding protein I	23-32	37-46	50-59	89-98
54_4	5252	fibronectin binding protein I	43-52	66-75	95-104	147-156
54_5	5253	fibronectin binding protein I	49-60	81-90		
54_6	5254	fibronectin binding protein I	55-71	82-97	139-158	175-186
328_1	5255	lipoprotein (H. flu)	11-20	61-70	96-105	

Table 4

ORF	Antigenic Regions (cont)					
	Region 5	Region 6	Region 7	Region 8	Region 9	Region 10
46_1	215-242	333-352	376-385	416-432	471-487	
63_4	145-154	191-202	212-223	245-265	274-283	291-300
174_6						
206_16	239-259	275-284				
267_1						
322_1	298-319	350-359				
415_2	344-353	371-380	395-404	456-465	486-495	518-527
214_3	318-337	365-375				
587_3	106-115	142-151	156-166	173-182	186-198	204-213
685_1	113-122	130-145				
54_3	128-138	185-194	217-226	251-260	268-277	295-305
54_4	175-188	191-200	203-212	220-229		
54_5						
54_6	220-230	287-304	317-326	344-353	364-373	378-387
328_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 11	Region 12	Region 13	Region 14	Region 15	Region 17
46_1						
63_4	306-315	319-328	366-376	395-420	453-462	467-476
174_6						
206_16						
267_1						
322_1						
415_2	539-555					
214_3						
587_3	217-226	278-287	318-327	332-342	351-360	377-386
685_1						
54_3	316-325	329-345	355-372	387-396	416-425	438-448
54_4						
54_5						
54_6	396-407	427-436	514-531	541-550	569-578	612-622
328_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 18	Region 19	Region 20	Region 21	Region 22	Region 23
46_1						
63_4	485-500	513-525				
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	396-405	426-442	459-470	485-494	505-514	531-562
685_1						
54_3	455-462	472-491	517-536			
54_4						
54_5						
54_6	639-648	673-681	703-715	723-732	749-760	772-788
328_1						

Table 4

ORF	Antigenic Regions (cont)					
	Region 24	Region 25	Region 26	Region 27	Region 28	Region 29
46_1						
63_4						
174_6						
206_16						
267_1						
322_1						
415_2						
214_3						
587_3	567-578	584-601	607-840	844-854	858-870	877-886
685_1						
54_3						
54_4						
54_5						
54_6	793-802	811-826	834-848	866-876	893-903	907-918
328_1						

Table 4

ORF	Antigenic Regions (cont)	
	Region 30	Region 31
46_1		
63_4		
174_6		
206_16		
267_1		
322_1		
415_2		
214_3		
587_3	889-911	927-936
685_1		
54_3		
54_4		
54_5		
54_6	925-944	951-997
328_1		

SEQUENCE LISTING

5

(1) GENERAL INFORMATION:

(i) APPLICANT:

10

- (A) NAME: Human Genome Sciences, Inc.
- (B) STREET: 9410 Key West Avenue
- (C) CITY: Rockville
- (D) STATE: Maryland
- (E) COUNTRY: US
- (F) POSTAL CODE: 20850

15

(ii) TITLE OF INVENTION: Staphylococcus aureus Poly-nucleotides and Sequences

20

(iii) NUMBER OF SEQUENCES: 5255

(v) COMPUTER READABLE FORM:

25

- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
- (B) COMPUTER: HP Vectra 486/33
- (C) OPERATING SYSTEM: MSDOS version 6.2
- (D) SOFTWARE: ASCII Text

30

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:

35

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 60/009,861
- (B) FILING DATE: 05-JAN-1996

40

(2) INFORMATION FOR SEQ ID NO:1:

45

50

55

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5895 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

10 TCCATTATGA AGTCACAAGT ACTATAAGCT GCGATGTTAC CAATGTTTTT TAAAATCCCA 60  
 GTAATAAAAT CAAAAAATAA GTTAAATAAT GTATTCAATT TAAGTCCTCC TTAATAAAGa 120  
 15 aaataGGTAA TAATGTAATA GCTTCTATTA TGATGCCTAA TTGAATGAAT TGGGCAAATG 180  
 GCTCTTTGAT GATAAGTGTG ATAATGAAAA GGGTTAAACT AACATAATC GCATAATATT 240  
 TTTTCGTTT AATAAGTCGC ACAGGAATGG GCTTCTTTTT AGTTGCTGCA GGAGCATATA 300  
 20 CTGAGATTAC ACCTAAAGAA ATAAGTGTAA AAATAATCAT AATTAAAAAG TTAATATGAA 360  
 AATTTACTAT TACTAAAGGT AAAAGTATAA ATAGTATAAT ACTTTCTACA TAACACCAAA 420  
 AAGAAGAAGG TGCATGTGCa CCATGTGCAT GtCTTCTTAT TAAATAAAAT GTTAAATTCTG 480  
 25 TAATTAACGT AAACAGAAAA ATGTTTAAAA TATAGGCAAT AGTATACATA ACAATTAATT 540  
 TACCTATATT TTTAGCTAAG ACCTGCATCC CTAATCGTAC TTGCAAAAAT TGAATATGAT 600  
 CTAAGTTATT TCTCTTTTGA AGATACGTGG CAACTGGGTC AATTTTATTA TCAAAATAAT 660  
 TCAATTTTAC ACCACTCTCC TCACTGTCAT TATACGATTT AGTACAATCT TTTATCATTa 720  
 TATTGCCTAA CTGTAGGAAA TAAATACTTA ACTGTTAAAT GTAATTTGTA TTTAATATTT 780  
 35 TAACATAAAA AAATTTACAG TTAAGAATAA AAAACGACTA GTTAAGAAAA ATTGGAAAAT 840  
 AAATGCTTTT AGCATGTTTT AATATAACTA GATCACAGAG ATGTGATGGA AAATAGTTGA 900  
 TGAGTTGTTT AATTTTAAGA ATTTTATCT TAATTAAGGA AGGAGTGATT TCAATGGCAC 960  
 40 AAGATATCAT TTCAACAATC GGTGACTTAG TAAAATGGAT TATCGACACA GTGAACAAAT 1020  
 TCACTAAAAA ATAAGATGAA TAATTAATTA CTTTCATTGT AAATTTGTTA TCTTCGTATA 1080  
 GTACTAAAAG TATGAGTTAT TAAGCCATCC CAACTTAATA ACCATGTAAA ATTAGCAAGT 1140  
 45 GAGTAACATT TGCTAGTAGA GTTAGTTTCC TTGGACTCAG TGCTATGTAT TTTTCTTAAT 1200  
 TATCATTACA GATAATTATT TCTAGCATGT AAGCTATCGT AAACAACATC GATTTATCAT 1260  
 50 TATTTGATAA ATAAAATTTT TTTTATAATT AATAACATCC CCAAAAATAG ATTGAAAAAA 1320  
 TAACTGTAAA ACATTCCCTT AATAATAAGT ATGGTCGTGA GCCCCTCCCA AGCTCGCGGC 1380  
 CTTTTTTGTA ATGAAGAAGG GATGAGTTAA TCATCATTAT GAGACCCGCC GTTAAAATAT 1440



	TCATTTGCAA	AGGGCGAAAT	GGGTTCTTAC	TGAGTTATCT	ATTATAAAAA	AATAAACATA	1560
	GACTTATGAA	AAATCTCTCA	TAAATCTATG	TTTAGTCATG	aCATGTGTTA	AATATTATTT	1620
5	CGGGCGCTTC	TTATTTATAC	AAATCTAATT	TAATACTTTT	AAATACAGGT	ATATTTTCgC	1680
	GTTGCTGTTC	TACTTCATTT	AAGTTTAAAT	CTACAGTCAA	AATATCTGCG	GATTCAATTTA	1740
	ATTCTCCAAC	TAAATCTCCA	TTTGGGTTTA	TAACATCGA	ATGACCAGCA	TATTCTGTGT	1800
10	TACCATCGAA	TCCAGTGCTA	TTAGTTCCAA	TGACAAACAT	ATTATTTTCA	ATTGCACGTG	1860
	CCTTTAGTAA	TGAATGCCAA	TGTTGAAGAC	GTGACATAGG	CCATTGCGCC	ACATAAAATG	1920
15	CAATTTTAGC	ACCACTACGA	GCAGGATATC	TTAATAATTC	TGAAAACGT	AAATCATAAC	1980
	AGATAAGTTG	GGTCACATAA	GTACCGTCAG	ACAATTGAAA	GGGTTGAGCT	ACGTATTTCGC	2040
	CAGCGGTAA	AAATTCATGC	TCTCTTAACA	TAGGAACTAA	ATGAACTTTG	TCGTATTCaT	2100
20	TAATCAGCTG	GCCACTTTTA	TTCACACTAA	AAGCTGTATT	AAATATTTGA	TTGTTTCTAA	2160
	TGTTAGAAAC	TGACCCAGCT	ACGATATCGA	CTTTATATTT	TTCAGCTAAA	TGTTTAATAA	2220
	ATGAAAAACT	TTGTCCTAGA	TTATTATCTG	CTTTTTTCATT	TAAATGCTCT	AAATCATAGC	2280
25	CATTATTCCA	CATTTCAGGT	AAAACGACTA	CATCTACTTC	AGCATTCTA	TTTTTTTCGA	2340
	ACCATTGCGT	TATTTGAGTT	TCATTTTTAG	AACTATCTCC	AAAAACAATC	GGTAATTGAT	2400
30	AAATTTGGAC	TTTCATAACA	TCACATCCTT	GATAGATCTT	ATATATAACT	TACTAAAAGT	2460
	TATGTTGAAA	CGCAAAAAAC	GAGCACAAGA	CATAAAATCA	AAGTCCTAGG	CTCTACAAAG	2520
	TTATATTGAC	AGTAGTTGAT	GGGGCCCCAA	CATAGAGAAA	TTGGAACACC	AATTTCTACA	2580
35	GACAATGCAA	GTTGGGGTGG	GCTCTAACAT	AAAGAAATAC	TTTTTCTTTA	GAAATTAGTA	2640
	TTTCTTATAC	ATGAGTTTTA	CTCATGTATT	CCTATTCTTA	AGTGCACATT	AGCAGCGGCT	2700
	AATGTGTAAG	AACTACTACA	TAATGAATAA	CTAATGATTC	TTTATCATTT	CTGTCCCATT	2760
40	CCTAACAATA	TATTGATTAT	TTTTTTATTA	CGAAACGATC	TTCCACTGGA	TTAAATGTTT	2820
	TTTCGCCAGC	AGCTTCACGA	ATATCACCAA	ATGGCATTTC	AGCAATAAGT	TTCCAACTTT	2880
45	TAGGAATATT	AAATTCATTT	GAAGTCATCT	CATCAACAAG	TGGATTATAG	TGTTGTAATG	2940
	AAGCACCTAT	GCCTTTAGTA	GCTAATGCAG	TCCAAATTGC	AAATTGATGC	ATGGCATTTC	3000
	TTTGAGTTGA	CCATATTGCA	AAATTATCAT	AGTAGTTTGG	CATTTGTTCT	TGTAAACCAC	3060
50	TTACAACATC	TTGATCTTCA	TAAAACAAAA	TTGTACCGTA	TGAATGTTTG	AAGTTATCAA	3120
	TTTTTTGTTC	AGTTGGCTCG	AAATCACGAT	TCTCTCCCAT	GACTTCTTTT	AAAATTGCTT	3180
55	TTGTGTTATC	CCAAAATTTA	TTATTGTTGT	CATTTAACAA	GAGAACAATT	CTAGTTGATT	3240

	CATCGCTAAT	TGATATCGAA	TCTTTCAAAT	TATATATTGA	ACGTCTTTCT	TCCATTGCAT	3360
	TGTCAAAAGT	CATTGCTTTT	TTATCTTTTT	TAAATAAGCC	CATAATTATT	GCTCCTTCTT	3420
5	TAGTAAAGAA	TACTTAATAG	ACTAAGTATA	AAATTTATAC	TCGTACTTGT	AAAGCAATAT	3480
	TTACGAAAAT	TTCAAGAATA	TTAATATTCA	TTTTCAAATT	CCAAATATAA	ATGCATTTTC	3540
	AACGCATATT	TATTATACTT	AGATTAATAC	TTACATGAAA	AAGGGAGGTG	TCTCGTGAAA	3600
10	TGTCATATCA	TTGGTTTAAG	AAAATGTTAC	TTTCAACAAG	TATTTTAATT	TTAAGTAGTA	3660
	GTAGTTTAGG	GCTTGCAACG	CACACAGTTG	AAGCAAAGGA	TAACTTAAAT	GGAGAAAAAC	3720
15	CAACTACTAA	TTTGAATCAT	AATATAACTT	CACCATCAGT	AAATAGTGAA	ATGAATAATA	3780
	ATGAGACTGG	GACACCTCAC	GAATCAAATC	AAACGGGTAA	TGAAGGAACA	GGTTCGAATA	3840
	GTCGTGATGC	TAATCCTGAT	TCGAATAATG	TGAAGCCAGA	CTCAAACAAC	CAAAACCCAA	3900
20	GTACAGATTC	AAAACCAGAC	CCAAATAACC	AAAACCTCAAG	TCCGAATCCT	AAACCAGATC	3960
	CAGATAACCC	GAAACCAAAA	CCGGATCCAA	AACCAGACCC	AGATAAACCA	AAGCCAAATC	4020
	CGGATCCAAA	ACCAGATCCA	GATAACCCGA	AACCAAATCC	AGATCCAAAA	CCAGACCCAG	4080
25	ATAAACCAAA	GCCAAATCCG	GATCCAAAAC	CAGATCCAGA	TAAACCAAAG	CCAAATCCGA	4140
	ATCCAAAACC	AGACCCTAAT	AAGCCAAATC	CTAACCCGTC	ACCAGATCCC	GATCAACCTG	4200
30	GGGATTCCAA	TCATTCTGGT	GGCTCGAAAA	ATGGGGGGAC	ATGGAACCCA	AATGCTTCAG	4260
	ATGGATCTAA	TCAAGGTCAA	TGGCAACCAA	ATGGGAATCA	AGGAAACTCA	CAAAATCCTA	4320
	CTGGTAATGA	TTTTGTATCC	CAACGATTTT	TAGCCTTGGC	AAATGGGGCT	TACAAGTATA	4380
35	ATCCGTATAT	TTTAAATCAA	ATTAATAAGT	TGGGCAAAGA	TTATGGAGAA	GTTACTGATG	4440
	AAGACATTTA	TAATATTATT	CGAAAACAAa	ATTTTCAGCG	AAATGCATAT	TTAAATGGAT	4500
	TACAACAGCA	ATCGAATTAC	TTTAGATTCC	aATATTTCAA	TCCATTGAAA	TCAGAAAGGT	4560
40	ACTATCGTAA	TTTAGATGAA	CAAGTACTCG	CATTAATTAC	TGGTGAAATT	GGATCAATGC	4620
	CAGATTTGAA	AAAGCCCGAA	GATAAGCCGG	ATTCAAAACA	ACGCTCATTT	GAACCGCATG	4680
45	AAAAAGACGA	TTTTACAGTA	GTTAAAAAAC	AAGAAGATAA	TAAGAAAAGT	GCGTCAACTG	4740
	CATATAGTAA	AAGTTGGCTA	GCAATTGTAT	GTTCTATGAT	GGTGGTATTT	TCAATCATGC	4800
	TATTCTTATT	TGTAAAGCGA	AATAAAAAGA	AAAATAAAAA	CGAATCACAG	CGACGATAAT	4860
50	CCGTGTGTGA	TTCTTTTTTT	TTATTATGGA	ATAAAAATGT	GATATATAAA	ATTGCTTGT	4920
	TCCGTGGCTT	TTTTCAAAGC	CTCAGGATTA	AGTAATTGGA	ATATAACGAC	AAATCCGTTT	4980
55	TGTAACATAT	GGATAATAAT	TGGAACAGCA	AGCCGTTTTG	TCCAAACATA	TGCTAATGAA	5040

AATATTAATG AACTTACTGT TG TAGCAATA ATAAATGCCA CGATACGATT ACCTTTAATC 5160  
 GCATTAAATA ATTCTCCAAA GATTACTTTT CTGAATACAT ATTCTTCTAA TAAAGGACCA 5220  
 5 ATAATAGATA CAAAGAAGAT AAATATAGGT ATTTTTCGAG CAATAATAAT TAGCTTTTCT 5280  
 GTATTAGGAC TTACTTGTTG TCCACCATAA ATTTGCGTTA ATACAATGCT CACTACCATT 5340  
 10 TGATAAATCA TTACCAATGC AAATCCAAGC AATGCCCATG GAATGATATA TTTTITAGGT 5400  
 TCTTTAACTT CTAATTCTAA TTTTGTTGGA TTTTAAATTT TTAAATTAAT TAAAATAATC 5460  
 GTCGTGGCGG CGATTAAAAA TAGAACAAGT TGTATGTAAA TGACTGCTTT AGTCAGTTCT 5520  
 15 ATGCCACTAT ATTGTACAAA TGGTAATTTT TTTACAATGA GAAGCGGTAA AAATTGAGAC 5580  
 AATATATAAA TAATAACAGT TAGCAATGAT GCCCATAATC tTGCATAAT TTTCTCCAA 5640  
 ATATTTGTTT ATAATTTATT TTATCGTAAA TAACTTGAAG TTACAAAAC TAATTAAAAG 5700  
 20 GTTATGACTT GAAATTTTGA CCAATTTTGA TTATTATAAA TGTATGTTAG CACTCTTTAA 5760  
 TGTTAAGTGC TAAACTTTAG GTTTTTTAAG GAGGAACAAT CATGCTAAAA CCAATTGGAA 5820  
 ATCGTGTGAT TATTGAGAAA AAAGAACAAG AACAAACAAC TAAAAGTGGn ATTGTTTAAC 5880  
 25 TGATAGTGCT AAAGA 5895

## (2) INFORMATION FOR SEQ ID NO: 2:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6796 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

TTTGAAAAAA CAAGGTACGA TTGGTTTAAT AACATATATG AGAACCGATT CTACACGTAT 60  
 40 TTCaGATACT GCCAAAGTTG AAGCAAAACA GTATATAACT GATAAATACG GTGAATCTTA 120  
 CACTTCTAAA CGTAAAGCAT CAGGGAAACA AGGTGACCaa GATGCCCATG AGGCTATTAG 180  
 45 ACCTTCAAGT ACTATGCGTA CGCCAGATGA TATGAAGTCA TTTTIGACGA AAGACCAATA 240  
 CCGATTATAC AAATTAATTT GGGAACGATT TGTGCTAGT CAAATGGCTC CAGCAATACT 300  
 TGATACAGTC TCATTAGACA TAACACAAGG TGACATTAaa TTTAGAGCGA ATGGTCAAAC 360  
 50 AATCAAGTTT AAAGGATTTA TGACACTTTA TGTAGAAACT AAAGATGATA GTGATAGCGA 420  
 AAAGGAAAAT AAAGTGCCTA AATTAGAGCA AGGTGATAAA GTCACAGCAA CTCAAATTGA 480  
 ACCAGCTCAA CACTATACAC AACCACCTCC AAGATATACT GAGGCGAGAT TAGTAAAAAC 540  
 55

	AAAGCGTAAC TATGTCAAAT TAGAAAGTAA GCGTTTGTGTT CCTACTGAGT TGGGAGAAAAT	660
	AGTTCATGAA CAAGTGAAAG AATACTTCCC AGAGATTATT GATGTGGAAT TCACAGTGAA	720
5	TATGGAAACG TTACTTGATA AGATTGCAGA AGGCGACATT ACATGGAGGA AAGTAATCGA	780
	CGGTTTCTTT AGTAGCTTTA AACAAAGATGT TGAACGTGCT GAAGAAGAGA TGGAAAAGAT	840
10	TGAAATCAAA GATGAGCCAG CCGGTGAAGA CTGTGAAATT TGTGGTTCTC CTATGGTTAT	900
	AAAAATGGGA CGCTATGGTA AGTTCATGGC TTGCTCAAAC TTCCCGGATT GTCGTAATAC	960
	AAAAGCGATA GTTAAGTCTA TTGGTGTTAA ATGTCCAAAA TGTAATGaTG GTGACGTCGT	1020
15	AGAAAGAAAA TCTAAAAAGA ATCGTGTCTT TTATGGATGT TCGAAATATC CTGAATGCCA	1080
	CTTTATCTCT TGGGATAAGC CGATTGGAAG AGATTGTCCA AAATGTAACC AATATCTTGT	1140
	TGAAAATAAA AAAGGCAAGA CAACACAAGT AATATGTTCA AATTGCGATT ATAAAGAGGC	1200
20	AGCGCAGAAA TAATATTTTT ATTTCTTAGA TACATTTTAA GATTGTTAAA TAGAATCATT	1260
	AGTGAATCTT ATTTTAAAGA TAGTAAAGGA TTAATCTAAA TAAGTGCGGA TAATATAAAC	1320
25	ATAACAACAT AATTAAAGAGA CATAAATGAC aATAAAAGGA GTATAGAAAT GACTCAAAC	1380
	GTAAATGTAA TAGGTGCTGG TCTTGCCGGT TCAGAAGCGG CATATCAATT AGCTGAAAGA	1440
	GGAATTAAAG TTAATCTAAT AGAGATGAGA CCTGTAAAC AAACACCAGC GCACCATACT	1500
30	GATAAATTTG CGGAACTTGT ATGTTCCAAT TCATTACGCG GAAATGCCTT AACTAATGGT	1560
	GTGGGTGTTT TAAAAGAAGA AATGAGAAGA TTGAATTCTA TAATTATTGA AGCGGCTGAT	1620
	AAGGCACGAG TTCCAGCTGG TGGTGCATTA GCAGTTGATA GACACGATTT TTCAGGTTAT	1680
35	ATTACTGAAA CACTTAAAAA TCATGAAAAT ATCACAGTTA TTAATGAAGA AATTAATGCC	1740
	ATTCCAGATG GATACACAAT TATCGCAACA GGACCACTTA CTACAGAAAC CCTTGCGCAA	1800
40	GAAATAGTGG ACATTACTGG TAAAGATCAA CTTTATTTCT ATGATGCGGC TGCTCCAATT	1860
	ATTGAAAAAG AATCTATTGA TATGGATAAA GTTTACTTAA AGTCCCGTTA TGATAAAGGT	1920
	GAAGCTGCAT ATTTAACTG TCCTATGACT GAGGATGAAT TTAATCGCTT TTATGATGCA	1980
45	GTATTAGAAG CTGAAGTTGC GCCTGTAAAT TCATTTGAAA AAGAAAAATA TTTCGAGGGT	2040
	TGTATGCCTT TTGAAGTAAT GGCAGAACGC GGACGCAAGA CATTACTATT TGGACCAATG	2100
	AAACCAGTAG GATTAGAAGA TCCAAAGACT GGGAAACGTC CTTATGCGGT GGTTC AATTA	2160
50	AGACAAGATG ACGCTGCTGG TACACTCTAC AATATTGTTG GCTTCCAAAC GCATTTAAAA	2220
	TGGGGAGCTC AAAAAGAAGT CATTAAATTA ATTCCAGGCT TAGAAAATGT TGATATTGTT	2280
55	AGATATGGTG TGATGCATAG AAATACCTTC ATTAATTCAC CGGACGTATT AAACGAGAAA	2340

	TATGTAGAAA GCGCagcTAG CGGCTTAGTT GCAGGTATCA ATCTTGCGCA TAAAATATTA	2460
	GGCAAGGGTG AGGTAGTATT TCCGAGAGAA ACAATGATTG GAAGTATGGC TTACTATATT	2520
5	TCTCATGCTA AAAACAATAA GAATTTCCAA CCTATGAATG CTAACCTCGG GTTATTACCA	2580
	TCTTTAGAAA CTAGAATTAA AGATAAAAAA GAACGCTATG AAGCACAAGC TAATAGAGCT	2640
10	TTGGATTACT TAGAAAATTT CAAAAAACT TTATAAAATA GTTAGAAAGA CTAGATATGC	2700
	TATTCATTCT TAAGTCATCA ACGAGTAAGT AATGACTTTC TAAATGGAAA ATACTTATCC	2760
	TAGTCTTTTT AATTTTGGAA TTGTTACGTA TTTCTGACAA TTTAGAATTC GCATTCAAAA	2820
15	AATATCTAAA TAAATAACAC GCAATAAGTT GATTGATGTA ACATGTAAGA GAATGTTTTA	2880
	AATAAACTTT ATTTAAAGG CAATGAAATA ATAAATGGCA AGGCTATTAA TAAAGACTTT	2940
	TAGTAATTAA TTAAAAAAG AGGTATTCTA ATTAACAGGT TTTCCGATTG GTTACAATTA	3000
20	TTTAATTCTC AAAAGATTTA GAATTGATTA TCAAATTACT GTAAGCCCTT TGCTGTATAT	3060
	GCTACAATTC TTATTGATGG AGGGTAAATG TATTGAATCA TATTCAAGAT GCGTTTTTAA	3120
25	ATACATTGAA AGTTGAACGG AATTTTTTCG AACACACATT GAAATCATAT CAAGATGACT	3180
	TAATTCAGTT TAATCAATTT TTAGAACAAG AACATTTAGA GTTGAATACT TTTGAATACA	3240
	GAGATGCTAG AAATTATTTG AGCTATTTAT ATTCAAATCA TTTGAAAAGA ACATCTGTTT	3300
30	CTCGTAAAAT CTCAACGTTA AGAACTTTCT ATGAATATTG GATGACGCTT GATGAGAACA	3360
	TTATTAATCC ATTTGTTCAA TTAGTACATC CGAAAAAGA AAAATATCTT CCGCAATTCT	3420
	TTTACGAAGA AGAAATGGAA GCGTTATTCA AACTGTAGA AGAGGACACT TCAAAAAATT	3480
35	TACGGGATCG AGTTATTCTT GAATTGTTGT ATGCTACAGG CATCCGTGTT TCGGAATTAG	3540
	TAAATATTAA AAAACAAGAT ATAGATTTTT ACGCGAATGG TGTTACCGTA TTAGGAAAAG	3600
40	GGAGCAAAGA GCGCTTTGTA CCGTTTGGTG CTTATTGTAG ACAAAGCATC GAAAATTATT	3660
	TAGAACATTT CAAACCAATT CAGTCATGCA ATCATGATTT TCTTATTGTA AATATGAAGG	3720
	GTGAAGCAAT CACTGAACGC GGTGTACGAT ATGTTTTAAA TGATATTGTT AAACGAACAG	3780
45	CAGGCGTAAG TGaGATTcAT CCCCACAAGC TCAGACATAC ATTTGCAACG CATTTATTGA	3840
	ATCAAGGTGC AGACCTAAGA ACAGTACAAT CGTTATTAGG TCATGTTAAT TTGTCAACAA	3900
	CTGGTAAATA TACACACGTA TCTAACCAAC AATTAAGAAA AGTGTATCTA AATGCACATC	3960
50	CTCGAGCGAA AAAGGAGAAT GAAACATGAG TAATACAACA TTACATGCAA CAACAATTTA	4020
	TGCTGTAAGA CATAATGGGA AAGCAGCTAT GGCTGGAGAT GGGCAAGTAA CGCTTGGTCA	4080
55	ACAAGTCATC ATGAAACAAA CGGCAAGAAA AGTGCGACGT TTATATGAAG GTAAAGTGTT	4140

	ATTACAACAG TTTAGTGGTA ACTTAGAAAAG AGCTGCTGTT GAATTGGCAC AAGAATGGCG	4260
	AGGCGATAAA CAATTACGTC AATTAGAAGC TATGCTAATT GTAATGGATA AAGATGCTAT	4320
5	TTTAGTTGTC AGTGGAAGTC GCGAAGTTAT TGCTCCAGAT GATGACCTTA TCGCTATTGG	4380
	ATCAGGAGGC AACTACGCAT TAAGCGCAGG ACGTGCATTG AAACGCCATG CATCGCATTT	4440
10	GTCTGCTGAA GAAATGGCAT ATGAGAGCTT GAAAGTAGCG GCTGATATTT GTGTCTTTAC	4500
	CAACGATAAT ATTGTTGTCG AAACACTATA ATAATCAGAG CACGATAAAT AATTACGAGC	4560
	AATTAATTTT AGTTAAAAGA CGGAGGAATG AAATTAATGG ATACAGCTGG AATAAGATTA	4620
15	ACTCCAAAAG AAATCGTATC TAAATTAAAT GAATACATCG TTGGACAAA TGATGCTAAA	4680
	CGTAAAGTGG CAATTGCCCT ACGTAATCGA TACAGAAGAA GTTTATTAGA TGAGGAATCA	4740
	AAGCAAGAAA TTTCACCTAA AAATATTTTG ATGATTGGAC CAACCGGCGT TGGTAAACT	4800
20	GAAATTGCAA GAAGAATGGC CAAAGTTGTC GGCGCGCCAT TTATAAAAGT AGAAGCTACT	4860
	AAATTTACTG AGGTAGGTTA TGTAGGACGA GATGTTGAAA GTATGGTTAG AGATCTTGTT	4920
25	GATGTTTCAG TAAGATTAGT CAAGGCGCAG AAAAAATCAT TGGTACAAGA TGAAGCAACA	4980
	GCTAAGGCCA ATGAAAAACT TGTTAAGTTA TTAGTTCCAA GTATGAAAA GAAAGCGTCT	5040
	CAAACGAATA ATCCTTTAGA GTCACTTTTT GGAGGTGCAA TTCCAAATTT CGGACAAAAT	5100
30	AACGAAGATG AAGAAGAACC ACCTACTGAG GAAATTAAAA CAAAACGTTT TGAAATTAAG	5160
	AGACAGCTAG AAGAAGGCAA ACTTGAAAAA GAAAAGGTAA GAATTAAAGT CGAACAAAGT	5220
	CCTGGTGCTT TAGGTATGCT AGGTACAAAT CAAAATCAGC AAATGCAAGA GATGATGAAT	5280
35	CAATTAATGC CTAAAAAGAA AGTTGAGCGA GAAGTTGCTG TTGAGACGGC AAGGAAAATC	5340
	TTAGCTGATA GTTATGCGGA TGAACATAAT GATCAAGAAA GCGCTAACCA AGAAGCGCTT	5400
40	GAATTAGCAG AACAAATGGG TATCATCTTT ATAGATGAAA TCGACAAAGT TGCGACGAAT	5460
	AATCATAATA GTGGTCAAGA TGTCTCAAGA CAAGGTGTTT AAAGAGATAT TTTACCTATA	5520
	CTTGAAGGTA GCGTTATTCA AACCAAATAT GGTACTGTGA ATACTGAACA TATGCTGTTT	5580
45	ATAGGTGCTG GAGCTTTCCA TGTATCTAAG CCGAGTGAAT TGATACCAGA ATTGCAAGGT	5640
	CGTTTTCCGA TTAGAGTTGA ACTTGATAGT TTATCGGTAG AAGATTTTGT AAGAATTTTG	5700
	ACAGAACCAA AATTGTCATT AATTAAACAA TATGAAGCAT TGCTTCAAAC AGAAGAAGTT	5760
50	ACTGTAAACT TTACCGATGA AGCAATTACT CGCTTAGCTG AGATTGCTTA TCAAGTAAAT	5820
	CAAGATACAG ACAACATTGG TGCACGTCGA CTTCATACAA TTTTAGAAAA GATGCTAGAA	5880
55	GATTTATCAT TCGAAGCACC AAGTATGCCG AATGCAGTTG TAGATATTAC CCCACAATAT	5940

AAATATACAA AAGGAGAAAA ATTCATGAGC TTATTATCTA AAACGAGAGA GTTAAACACG 6060  
 TTACTTCAAA AACACAAAGG TATTGCGGTT GATTTTAAAG ATGTAGCACA AACGATTAGT 6120  
 5 AGCGTAACTG TAACAAATGT ATTTATTGTA TCGCGTCGAG GTAAAATTTT AGGATCGAGT 6180  
 CTAAATGAAT TATTAAAAAG TCAAAGAATT ATTCAAATGT TGGAAGAAAG ACATATTCCA 6240  
 10 AGTGAATATA CAGAACGATT AATGGAAGTT AAACAAACAG AATCAAATAT TGATATCGAC 6300  
 AATGTATTAA CAGTATTCCC ACCTGAAAAC AGAGAATTAT TCATAGATAG TCGTACAACT 6360  
 ATCTTCCCAA TTTTAGGTGG AGGGGAAAGA TTAGGTACAT TAGTACTTGG TCnAGTACAT 6420  
 15 GATGATTTTA ATGAAAATGA TTTGGTACTA GGTGAATATG CTGCTACAGT TATTGGTATG 6480  
 GAAaTCTTAC GTGAGAAGCA TAGTGAAGTA GAAAnAGAAG CGCGCGATAA AGCTGCTATT 6540  
 ACAATGGCAA TTAATTCATT ATCTTATTCT GAAAAAGAAG CGATTGAACA TATCTTTGAA 6600  
 20 GAACTTGGCG GTACGGAAGG CCTATTAATC GCATCAAAAG TTGCAGATAG AGTTGGTATT 6660  
 ACTAGATCTG TAATTGTAAA TGCACTACGT AAATTAGAAA GTGCTGGTGT AATTGAATCA 6720  
 CGTTCTTTAG GAATGAAAGG TACTTTCATT AAAGTTAAAA AAGAAAAATT CTTAGATGAA 6780  
 25 TTAGAAAAAA GTAAAT 6796

## (2) INFORMATION FOR SEQ ID NO: 3:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2073 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

40 ATCCtAAAAT TnAAAAATTAT CACGCCTTTT GaACAGCTTT GTAACCaTct GGACGATCAT 60  
 KAAATTCCaA TGTAATCCT GGTTTAAaGT TGATCTTTAA CCTTATTTAA AycACCAATT 120  
 GTACGTATAT TATGTTGTTT AGCAAAATCA CGTTTTACAG CTAAAGCATA CGTATTGTTA 180  
 45 TACTTCATTG GTTTTAACAT AGTCATTTGA TATTTCTTTT CAAGACTTTG CTTAGCTTGT 240  
 TCATAAACTT TTTTCTCTTC TTTTGACTTC AATGGTTCTT TTGTTAATTC ACCTAAAACT 300  
 GTTCCAGTAA ATTCTAAATA CCCATCTATA TCGTCAGATT TTAAAGCATT AAATAAAAAAT 360  
 50 GCTGTTTTGC CCATACCATC TTTCACCTCT ACAGTATTTT TGGTCTCTTC TTCTATTAAA 420  
 ATTTTATACA TATTTGTAAT AATCGATGGC TCGGAGCCAA GCTTTCCAGC TAACGTAATT 480  
 55 TTATCACCTT TTTGTGCAAA CATAGGAATA GCGATAGCCA GTATAATAAT CATCACTATA 540

TCAAATATAA TTGCCAATAA GGCTGCTGGA ATTGCACCTA ATAATATCAA CGATGCATTG 660  
 TTACGGTCTA TACCTAATAA AATTAAATCT CCTAGTCCGC CTGCACCAAT TAATGCTGCT 720  
 5 AGTGTGCTG TACCTATAAT TAATACCATA GCCGTTCTTA CACCAGCCAT TATAACAGGC 780  
 ATTGCTATCG GAAGTTCGAC TTTAGTTAAA CGTCTAAATG GTTTCATACC TATACCTTTA 840  
 GCCGCTTCAA TGAGTGATGG ATCAACTTCT TTAATTCAG TATACGTATT CCTTAAATTT 900  
 10 GGTAACAACG CATACACTAC AAGTGCAATA ATTGCTGGCA CACGACCGAT ACCAAATAAA 960  
 GGAATCATT AACTAATAA TGCCAACGAT GGTATGGTTT GAAGAATTGC CGCAATATTC 1020  
 15 ATTACGATTT CAGATATCGT TTTAGTCTTC GTTAATAAAA TACCTAATGG TACCGCAATA 1080  
 GCAGTTGCAA TCAATAATGC GATAAATGAT ATTTGAATAT GTTCTATCAT TGTCGAAAAG 1140  
 AGTTGCCCT TACGTTCACT CAATATGTCg AAAAAGTTAG TCATGTTGAG CTACCTCCTT 1200  
 20 TTTCTGGGAC AAATATTGA AGATATCTTT CCTATCAATA ACATATTGAC CTACGCTATC 1260  
 TTCTTGCATG ACAATGACAC GCTCGCTCTC TGATAAAAGT TGATACAATA CTTCAATTGG 1320  
 TTGATTGTCA TAAACAATTG GATAAGCGCT CATAGATGTA ACCTCATCGA TTGGTTTCAT 1380  
 25 AATATCCAAG TCACGGATAA TTGCGTTCTC TTCAACACAT GGCGCATCAT CTTCTAAATG 1440  
 ACTACCCATA AATTGTTTAA CAAATTCACT TTGAGGATTA TTTTAAATC CTTCTGGTGT 1500  
 GTCAATTTGT TCAATATGCC CTTCAATCAA AAGACAAATC TTATCACCAA GTTTCATCGC 1560  
 30 CTCTTGAATA TCATGTGTAA CAAATATGAT TGTCTTCTTA ATTTTAGTTT GTAATTCAAT 1620  
 TAAATCATCT TGAAGTTTTT CTCGGCTGAT TGGGTCTAAT GCACTAAACG GTTCATCCAT 1680  
 35 TAAAATAACT GGTGGATCAG CTGCTAACGC ACGTATAACT CCTACACGTT GTCGTTGCCC 1740  
 CCCTGACAAT TCATCAGGTT TTCTGTTTTT ATATTTTTCA GGTTCTAATC CAACCATTTC 1800  
 AAGTAATTCA TCTACTCTT TATCTATATC TTTTCTTTC CACTTTTTCA TTTGTGGCAC 1860  
 40 TTGTGCAAtA TTTTCTTTGa WTGTCaTATG TGGAATAAT GCAATCTGCT GcAATACGTA 1920  
 TCCAATATCC CAACKCATTT CGTATACTGG ATAATCACTT ATTGGTTTAT CTTTAAATA 1980  
 45 AATATAACCT TCACTTAAAGT GAATGAGTCG ATTAATCATT TTTAATGTCG TAGTTTTTCC 2040  
 ACAACCTGAA GGTCCAATTA GCACAAAAAA TTC 2073

(2) INFORMATION FOR SEQ ID NO: 4:

- 50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13321 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 55



EP 0 786 519 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

	ACTATTCTAG CTTCATCAGT TATCATATAT TCTTTGAAAC ACTTGTAAGA AAATATAATG	60
5	AGTATTTACT ACATAATGAT ATTTCAAATT AGAAAAAAGG AAGTTATGAT TTAATGGCCT	120
	TGAGCCTATC ATAACCTCCT TTTATCATTT TATTGTTGTG TTGATGTTTC GATAACGTGG	180
	TACATCTTAT CAAACATCAA TTCGAAACCA TGCACCATGG CATCATGATA TTCTTTTTTC	240
10	TTTTGCTTGT ATTCTAAATT AGTAAATCGT CTTTCTTTTT CAACTAATGA ACGATAATAA	300
	AATAGCATTT GGGTGCCACC TGTTCACGT TCAAAAAATT CTACCTCAAT GACATCTTGC	360
15	GTTTCACTTA GTCCAGGCAT ACCGATAGTC ATCTTAACGT ATTCATCCAT AACTAAAGAT	420
	TCATAAATGC CTTCAATCAC ATTTACTTTG CCATTACGTT GTTGATCTAC AATACGATAT	480
	TTACCGCCTT CTTTAACGTC CGCTTCAATC TCTTTATTCTG TTCTGGCTGA TGTCATAAAC	540
20	CATTGTTTCA ACAAATCTTT CTTGTCCAA GCTTCGTATA CTAACCTCTGG AGAAAAATTA	600
	TAAAGCTTTT CAATTTCAAC TTCGACATGT TCATTCTCTA CATTAAATTT TGCCACTGTT	660
	GTCCACCCAC TTTCGCTCTT ACTTTTATTT TAACGTATTT TTGCTCAGTT CCAAACATAG	720
25	ATGATCATCA TTTTAAAAG ATTAGCGTTA TACGGTGAGT ACAACATGAT CTGTTAATAT	780
	AACAAGCCAC CTTACTTGGC TACATCGATA TATTGTTAAG CATTAAATGTT TCATTTCTTG	840
30	ACTAGTGTTT TTTTITAGCT TTGGAAAATT AAATAAAATC GCAATAAGTC CGCATAACCC	900
	TAATAATATA GGATAAATGC TGTATGGGAA TAACATTAAC GGTGAAATAC CAGCTACACC	960
	AGCCGCTGaA ATGACTTGCG GGCTATATGG TAATAAACCT TGGAAGCAGC CTCCAAATAT	1020
35	ATCAAGAATA CTGCTGATT TCCTTGAATC TACATCATAT TCATCTGCAA TATTTTITAGC	1080
	TAAAGGACCT GACATAATAA TAGAGATGGT GTTGTITGCC GTGGCAATAT CTGCGACACT	1140
	TACCAAACTA GCAATTCCTA ATTCTGCGCC ACGCTTTGAT TTCACTTTAG AGCGAACAAA	1200
40	TTGCAACAAC CATTCAATAC CACCATTGTG TTGAATAATA CCGACTAAAC CACCAATTAG	1260
	CAACGCAATC ATAGCAATAT CTTCCATGCT TATAATACCT TTGGACACTG CATCTAGTAG	1320
45	CCCCATCCAA CCGAATGAAC CATCTATGAG ACCAATGATT CCGGCTAATA ATGTTCCGCC	1380
	AATCAATACG ATAATGACAT TTACACCTAA TAATGCTAAT ACCAATACTA AGATATACGG	1440
	TACAACTTTA ATTAGATTAT AATCATAGTt TTTAGCATGA TTTAAAGAAA TGCCATTTCGT	1500
50	TAAGAAATAC AGAATAATAA TCGTTAAAAT AGCACCTGGC AATACAATTT TAAAGTTTAC	1560
	TCTGAATTTA TCTTTCATTT TCGTATGTTG TGTCTAACC GCAGCAATTG TTGTATCTGA	1620
55	AATCATTGAT AGATTATCGC CGAACATTGC ACCTCCAACA ACTGTAGCCa tTGctAGCGC	1680

	TCCTACAGAC	GTCCCATAG	ATATAGAAAC	AAACATACAA	ATCACAAACA	ATCCTACAAT	1800
	AATTAAATTT	TCTGGGATTA	ATGATAGTCC	TAAATTAAC	GTCGACTTTA	CGCCACCCAT	1860
5	TTTTTCAGCT	GTATTTGAAA	ATGCACCTGC	TAAAATAAAA	ATCAACATCA	TTAAAACAAT	1920
	GTTTGAATGG	CCTGCACCTT	TCGTGAAGAC	CTCAACTTTT	TTAGCAAATG	ATTCTTTTCG	1980
10	ATTCATTAAT	AACGCCACAA	TTACCGTTAT	CGTAATTGCA	ACATTTAATG	GCATTGAAGT	2040
	AAAATCACCT	GTGATAATAC	CTACGCCTAA	AAACAACGCC	ACAAATAATA	ACAAGGGGAA	2100
	TAATGCCCAA	GCATTGCTCT	TTTTATGTAC	TTCCATCCTT	TTTACCTGCT	TTCCAATTAA	2160
15	AAATACCTCT	TTCTCACAAA	CGATGAAGAA	AGAGGTTTTT	ATGTGCTTTA	CCTGCTTATC	2220
	TTCAAACCAT	TACGGTACT	GGAATTGGCA	CATTTCGAGAT	GTTGCCGAGG	CTTCATAGGG	2280
	CCAGTCCCTC	CACCTCTCTA	GATAAGTGAT	GCTTATTTAC	GTTTACGTTA	CAAGATAATC	2340
20	CTTAGTACGT	CAATCATAAA	TTAATCAGGA	GTCGTATAAT	ATTTTTCATA	AACAATCATT	2400
	GCTACTGTAA	TAATAATCAA	AACAATAATG	CTAATAACAA	GTAAAAGCCA	CCATTTAAGC	2460
25	ATTAATGCAA	TAAAAATGAA	CACGATAGAC	ACACTTACTA	ATATTAATGA	TATGACTTTA	2520
	AATTGCTGAA	CACGTTGCTT	GGAGATGACT	TTCAACTGTT	TGTTTGATAG	ACGCGTATTT	2580
	TTTATACTGA	TTCCCAGTAT	ATTTTCTAAT	ATTTGAACCA	ATACGATACT	TATTGCAAAT	2640
30	ATAATAATTG	GTAAAACATC	ATAGCTCCCT	ATAGTTAATG	TATAAATTAC	AAATCCAATG	2700
	TAAAGTAACC	CTGAGACAAA	GGATAAAAAG	TATGCGACGT	ATTTGTATAA	CTTAATGATA	2760
	TGCTTTTTAA	CGTTTTGATG	TGTAAACCAT	ACATTCGAAA	CGATCGCAAC	TGCTACAAAT	2820
35	AATGTGAATA	CTATATATAA	TGGTAATTTT	TGTTTCAGGAA	AAACAGTCGC	TATTCCAAAA	2880
	GCTAATGCTA	AAATCAAAAA	TAATATAGCT	CTAGATACTA	TTAATGCCAT	AATAACAACC	2940
40	CCTTTGTTTA	ATATCGAGTT	TGCAAATTTA	CGTTTATCAG	CGTTTCTATG	ATCAGTACTT	3000
	CTACGGGTAG	CGTTTCTATG	TAATTTACAT	CATCTTAACA	TATAAATACT	TCGCTATTTA	3060
	ATTGAAAACA	TATCCTATTA	TTCTTTGTCC	GTTCTGACGT	TTAATATCTA	GCCTTAGGCA	3120
45	TTTCACTTGT	TAATGAATTT	AACTTTCTTC	CACTAACCGT	CCCTAAACCC	AATCCCGCAA	3180
	CAGTTTTTAA	CTTTTTCGTT	GTTGTCCTGA	CATCCTCATT	AAGAAAGTTT	ATTCTGCTTA	3240
	AAACTTATAA	TCCACACCCT	GAGCAAACGC	TCCTTATGAC	AGAGTATTAA	AATAAGCCGA	3300
50	TAAAGATACA	CACCTTTACC	GACTATTTAA	AATACACTTC	ACCAATTCAT	TTTAATTTAA	3360
	TGGATTGAAG	TAACATAAAT	AATATTATGT	TGTTCAATTA	AAAGCTTCAT	ACAAACCTAA	3420
55	TCTATTTGCA	CTCCACCGCT	AACACCGAAC	ACTTGTCCGG	TTGTATAACT	TGATTCTTCT	3480

EP 0 786 519 A2

	GTGTTTGTGAC CAAATGTTGG GATTTTACTT TGAGGTTGTC CACCAGAAAT TTGTAATGGT	3600
	GACCAGAATG GACCAGGCGC TACACAGTTC ACTCTAATTC CTTTTGGTCC TAATTCCTCT	3660
5	GAAAACTTT TAGTTAATGA AATAATTGCT GCTTTTGAAG CGGCATAATC ATGAAGAATA	3720
	GGACTAGGAT TATAACCTTG TACAGATGAT GTCGTTGTAA TTGACGCACC CGGTTTTTAAA	3780
	TATTCGAATG CTTTTTGAAC TGTCCAAAAT AGCGGATAGA CATTGTTTC AAATGTTTCT	3840
10	GTAAATGCCT CAGTTGTAAA TCCATGAATA TCATCATGAT ACTGTTGATG TCCAGCAACT	3900
	AAAGTAACAT TATCTAAGCC ACCTAATTGT TGATATGCTT GTTCAACAAG GTCATAGTTG	3960
15	AACTGTTTCAT CTCTTATATC ACCAGGAATT AACACTGCCT TTTGACCACT TTCTTCAATC	4020
	ACTTGGCGTA CTTCTTGTGC ATCTTGTCT TCACTCGGAA GATAGTTAAT CGCTACATCT	4080
	GCACCTTCTT TAGCATACGC AATTGCTGCT GCACGCCCTA TTGCTGAGTC ACCACCTGTG	4140
20	ACTAATATTT TATAGCCTTG TAAGCGTTGA TGACCTTGGT AAGACGTTTC GCCACAATCG	4200
	GGTGCTGGCG TCATTTGAGA TTGTAAACCC GGTACCTCTT GTTCTTGTTT TTCATAATCC	4260
	GTTGTTTTAA ATTTTGTCT AGGATCTTGA GCTGCCATTT TTTTACATCT CCTTATTCGC	4320
25	TTAATGGTTA TTATTTACCC AATCTTCCCTA GGAAGTTAAT CATGATTACA CTAAAAATTA	4380
	CTTCTTCTT TATAAAAACA AGCTCGAATT ATTCATGCAA TAGTCTCTTT ACAAATTCAA	4440
30	CAAAATACTC AGGTACTTTT TCCAGAATCC TTTTCATCCG TTTATATTGA GGATGATGTA	4500
	AATCATATTC ACTATGAGAA CCAATTAAAG CAAATACACT TGGAAAATGT TGACTATAAC	4560
	CTGAAAAATC TTCTCCAATC GTAAGCGGCT GTTCCATCAT TCCCACCTTA TATCCAACAT	4620
35	GTTGGGCTAC TGCAATTGCT TTATGCGTCA ATGCCTCATC ATTCATCACA GCGCCAGGTA	4680
	AATGCGTATA ATTTAAATTA ATTTTCATAT TATATGCTTG AGCCAATCCG TCCGCAATAT	4740
	CTTGJAATCG TGTTTCTACA AGCTTTCGTA CCACAGGATC AAAACTACGC ACTGTGCCTT	4800
40	GTACATACGC ATGATCAGCA ATGACATTCC AAGTATTACC ACATGATATT TGTCCAATTG	4860
	TTACTACCGC TTCATCAAAC GCAGATAGAT TTCTACTAAC TATGGATTGA ATACTATTAA	4920
45	TCAATTGCGC CAACACAATA ACTGGATCGT TGCATTGTTT TGGCTTTGCA GCATGACCAC	4980
	CCACGCCTTT AATATGAAAC TCAAAACGAT CTACTGCTGA TGTAATTGCC CCTGTTTTGA	5040
	TTGCAAATGT ACCTACCGAA CGCGATGGGT CATTATGAAA ACCCAATACT GCTTGACAT	5100
50	CTTTTAATGC ATGTGTTTCA ATAATTTTAA AAGCGCCATG TCCTAGTTCT TCTGCTGATT	5160
	GAAAAATGAA TTTAACACGC CCAGTAAGAG TGCCCTCAAT TTCTTTTAAT TTTACAGCTG	5220
55	TAGCCAAAAT ACTAGCCATG TGAATATCAT GACCACACGC ATGCATAACA CCTTCATTTT	5280

	CAGCTATACA	ACTCAGACCT	TGTCCCACTT	CAGCAACAAG	CCCAGTCGCA	AGTGGTAAGT	5400
	CTAATATTCT	AATATGATGT	TCTGTAAAA	TATCTTTAAT	TTTTTGTGTA	GTCTTAAATT	5460
5	CTTTATCGGA	TAGTTCTGGA	AATTGATGAA	AATACCTTCT	CCAGGTAACA	GCTTGATCTT	5520
	TTAATCCCAT	CGGTCATTCC	CCTTCCTTAA	GTCAATGATA	TGTTGTCTAC	CCTACGATGA	5580
10	TCATCTTTGA	CTATTAAACG	ATGATTTTAC	AACAATGTAC	TCTTGTTAAT	TGCTTTCGTT	5640
	AATGATAGAC	AGTTGTTTAA	TAATATCGTA	ACACTGTTGT	CAAACATTTC	TAACTTTTAT	5700
	AATTGAGACT	CTATACAAAA	ACGTGTTCTC	GAATATACTT	GTTTTTACAA	ACCACAAAAA	5760
15	GCTCTAAACA	TTAGTTTAAA	CCAATGCTTA	GAGCTTTCTA	ATTATTTTAT	GCTTTTAAAG	5820
	ATACTGTGTT	ATCTACGATG	ACCTTACCGT	CTTTAATAAC	TTTTTCTGCG	TGATTGATAC	5880
	CAAAATGATA	TGGAATATAT	TCATGATTTG	GTGCATCCCA	AATTACTAAA	TTAGCCTTAT	5940
20	CACCTGTGTT	AATTGTACCC	GCGTTAATGT	CTATTGCTTT	AGCAGCATTG	ACCGTAACAG	6000
	CATTCCAAAC	TTCATTAGGT	GATAGCTTTA	ATTTCAAGGC	TGCAATCGCC	ATAACAAGTT	6060
25	GTAAGTTGTT	TGTGACACTA	CTACCAGGGT	TATAATCAGT	TGCTAATGCA	ATCGCACCGT	6120
	TATTGTCAAG	CATGCCTCTT	GCATCTGCAT	AATCTTCTTT	ACCTAAATAG	AACGTCGTTG	6180
	CAGGTAAGAG	GACAGCTACA	GTATCACTAT	TTGCAACTT	TTCTTTTCCT	TTATCACTAG	6240
30	AAGCTACTAA	GTGGTCTGCT	GATATTGCTT	GTTCAATCAAT	TGCTAATTCC	AGTCCGCCTA	6300
	ACGGATCAAT	TTCATCCGCA	TGTATTTTCA	CTTTAAACC	TGCTTCTTTG	GCTTTTGTGA	6360
	TATAATGTTG	CGATTGTTCT	ATTGTAAATA	CACCTGTTTC	ACAGAAAATA	TCCGCAAAGT	6420
35	CTGCATATTG	TTTTACTTCC	GGAAGTAACG	CAATCATTTT	TTCTAAAAAT	GCCTCATTTG	6480
	AACTTGCCTC	TTTAGGTACA	GCATGAGGCC	CTAGGAAAGT	ATGTTTCATG	TCTAAATCAT	6540
	ATTCTCAGC	TAAACGATTA	GACACTTTCA	ATTGCTTCAG	TTCATTTTCT	CTATCTAATC	6600
40	CATAACCACT	CTTACTTTCA	ACTGCAAGCA	CGCCGTGTTT	AATCATAGTA	AGCAAATCAT	6660
	GCTCTGCTTT	TTTAAACAAG	TCATCTTCGG	ATGTTTCTCT	AGTAGCATTG	ACGGTAGATA	6720
45	ATATGCCACC	ACCCATTTCT	AATATTTCAA	GGTAAGACTT	ACCTTGACGT	TTTAATGACA	6780
	TCTCATGTTT	TCGAGATCCA	CCAAATGTTA	AATGGGTATG	TGCATCTACT	AATGCTGGGG	6840
	AACTACCTT	CCCACTAGCA	TCAATCGTCT	CAGTCGCATC	GATGTCATCT	GTATGTGTTT	6900
50	CAGCATATAC	AATTTTGCCA	TCTTTAATGA	CAACTGTACC	ATTTTTCACA	ACATTTAATT	6960
	CATCTAATTC	CTTACCCTTC	AAAGGTTTAT	CTGTTGATCT	CGGTAAATTT	AATTCTGCTA	7020
55	TATGATTAAT	TATTAAATCA	TTCATTACTT	ATCACCTGCT	TTATCAATCA	TTGGAATATG	7080

	AACACCCATA CCTGGGTCAG TCGTCAATAC ACGTTCCAAT CTTCTTTCAG CACGCTCTGA	7200
	TCCATCTGCT ACAACAACCA TACCCGCATG AAGTGAATAT CCCATGCCAA CACCGCCACC	7260
5	GTGATGGAAT GAAATCCATG AACCACCTGC AGCTGTGTTA ATGAGTGCAT TCAATACAGC	7320
	CCAATCACCA ACCGCGTCAC TACCATCTTT CATACTTTCT GTTTCACGGT TAGGACTAGC	7380
10	AACTGAACCA GCATCTAAAT GGTCTCGTCC AATAACAATT GGTGCTGAAA TTTCACCGTC	7440
	ACGTACAAGA CGATTTAAAG CTAAGCCCAT TTTCGCTCTT TCTCCATAGC CTAACCAAGC	7500
	AATACGTGAT GGTAGTCCTT GATATGAAAT TTTTCTTCA GCTAAATCAA GCCATCTTAA	7560
15	TAACTTTTCA TTTTCTGGGA AAAGTTTGGC CATTTCTTCA TCCGCACGCT CGATATCTTT	7620
	TGGATCACCA CTCAACGCAG CAAAGCGGAA TGGCCCTTTA CCTTCACAGA ATAATGGTCT	7680
	AATGTAAGCT GGTACAAAGC CTGGGAAGTC AAAAGCATTT TCACTCCGT TAITGAAGGC	7740
20	TACTTGACGA ATATTGTTAC CATAATCAAA TGCTACAGCG CCACGTTTTT GGAATTCAAG	7800
	CATTAATTCA ACATGCTTTG CCATTGAAGC TTGTGACAGT TCAACATATT TTTTCGGATC	7860
25	TTTTTCACGC AATACTTTG CTTCTTCTAC AGAGTATCCT TGTGGCACAT ATCCATTTAG	7920
	CGGATCATGT GCACTTGTTT GGTCAAGTAAT AATGTCAATT TTAAATCCTT TTTCTAGAAT	7980
	CGCTTGATGG ATGTCTACAG CATTTCCAAC TAACCCGATT GATAATCCTT CTCCACGTTT	8040
30	TTTCGCCTCT TCTGCTAATT TTAATGCTTC ATCTAAATCA GCTGTTTTAA CATCACAGTA	8100
	TTTCGTATCA ATTGCTTAT CAACACGTGT TTCATCAACA TCCACGCAA TTGCTACCCC	8160
	ATGATTCATA GTAATTGCTA ACGGTTGCGC ACCACCCATA CCACCTAAAC CTGCTGTCAG	8220
35	TGTAACAGTG CCTGCTAAAT CTCCATTAAA GTGTTGATTA CCTAGCTCGG CAAATGTCTC	8280
	ATAAGTACCT TGCACAATAC CTTGAGAACC AATATATATC CAACTACCGG CTGTCATCTG	8340
	TCCATACATG ATTAAACCTT TTTTATCTAA TTCAITAAAA TGATCCCAGT TTGCCCATTC	8400
40	AGGCACTAAT ACTGAATTTG AAATTAATAC ACGTGGCGCT TCTTCATGTG TTTTAAATAC	8460
	AGCAACTGGC TTTCCTGATT GTACTAACAT TGTCTCATCT GATTCTAATT CTCGTAACGT	8520
45	TTTCTCTATT GCTTCAAAAG CTTCCCAATT ACGTGCTGCT TTTCCAATAC CACCATAAAC	8580
	AACTAAATCT TCTGGTCTTT CAGCAACTTC TGGGTCTAAA TTGTTGTATA ACATTCTAAG	8640
	TACTGCTTCT TGTTCCCAAC CTTTACACTC AATACTCAAA CCTTTTTTTG CTTGAATTTT	8700
50	TCTCATAAAA TTCGCTCCTG TTCTTTTAAG AAGTTAATTC CACTAAATTT AAAACGCTTA	8760
	CATTATTATC TTCAATATTC ATTATAGTAT GTTAAATAT AGCCAACAAA TATAAATAAA	8820
55	CTAATTATCC ATAGCTTGAA TCTATAAATA AAAGGAGCAA AACACATGAA AATTATTCAG	8880

	CATATTAGCC	AGCCATCTTT	AACTGCTACG	ATTAAAAAAA	TGGAAGCAGA	TTTAGGTTAT	9000
5	GACTTATTTA	CACGTTCAAC	AAAAGACATC	AAGATTACCG	AAAAAGGAAT	ACAGTTTTAT	9060
	CGTTATGCGA	GCGAATTAGT	TCAACAATAT	CGATCCACGA	TGGAAAAAAT	GTATGATTTA	9120
	AGCGTTACAT	CAGAACCAAG	GATAAAAATT	GGGACTCTTG	AATCTACGAA	TCAATGGATT	9180
10	GCGAATTTAA	TTCGAAAGCA	CCATTCCGAC	TACCCTGAAC	AGCAATATCG	TTTATATGAA	9240
	ATACATGATA	AACATCAATC	TATAGAGCAA	TTACTGAATT	TTAATATTCA	TTTAGCTATA	9300
	ACAAATGAAA	AAATAACCCA	CGAAGATATA	AGATCCATTC	CTTTATATGA	GGAATCTTAC	9360
15	ATTTTATTAG	CACCCAAGGA	AACATTTAAA	AATCAAAATT	GGGTAGATGT	TGAAAATTTG	9420
	CCACTCATAT	TACCAAACAA	AAATTCTCAA	GTGCGCAAAC	ACTTAGATGA	CTATTTTAAT	9480
	AGAAGAAATA	TTCGTCCAAA	TGTCGTTGTA	GAAACAGATC	GATTCGAATC	AGCAGTTGGA	9540
20	TTTGTTTCATC	TCGGCTTAGG	TTACGCTATC	ATTCCGAGAT	TTTATTACCA	ATCATTTTAC	9600
	ACGTCTAATT	TAGAATATAA	AAAAATTCGT	CCAAACCTAG	GCCGAAAAAT	TTATATCAAT	9660
25	TACCATAAAA	AACGCAAACA	CTCCGAACAA	GTACATACAT	TCGTACAACA	ATGCCAAGAT	9720
	TATTTATATG	GACTTTTAGA	GGCTCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	9780
	CTCAGTCAAC	TGTATACCTT	TTGCCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	9840
30	CTCAGTCAAC	TGTATACCTT	TTGCCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	9900
	CTCAGTCAAC	TGTATACCTT	TTTCCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	9960
	CTCAGTCAAC	TGTATACCTT	TTGCCTTTAA	CTTAAGTTAT	TAGTGCCTCT	TATGTAGTTG	10020
35	CGTAGTCAaC	TGTaTACCTT	TTGCCTTTAA	CTTAAGTTAT	TAGAGCCTCT	TATGCAGTTG	10080
	CGCAGATCAT	CGTATAAAAA	TTAATGACGT	CATTTCAAAA	ATCGATACAA	AAATAATTTA	10140
40	TTATAAAAAAT	TCTAAGAAAG	AAGTGAAGCA	GATGTTAAAA	TCTATTAATC	ATATATGCTT	10200
	TTCAGTCAGA	AATTTAAACG	ATTCAATACA	TTTTTATAGA	GATATTTTAC	TTGGGAAATT	10260
	GCTATTGACT	GGTAAAAAAA	CTGCTTATTT	TGAGCTTGCA	GGCCTATGGA	TTGCTTTAAA	10320
45	TGAAGAAAAA	GATATACCAC	GTAATGAAAT	TCACTTTTCA	TATACACATA	TAGCTTTCAC	10380
	TATAGATGAC	AGCGAATTTA	AATATTGGCA	TCAGAGGTTA	AAAGATAATA	ACGTGAATAT	10440
	TTTAGAAGGA	AGAGTTAGAG	ATATTAGAGA	TAGACAATCA	ATTTACTTTA	CCGACCCTGA	10500
50	TGGTCATAAG	CTAGAATTAC	ATACTGGCAC	ACTTGAGAAC	AGATTAAATT	ATTATAAAGA	10560
	GGCTAAACCA	CATATGACAT	TTTACAAATA	AGGTGTCAAT	ATAAAAAGGC	CTCTTGAAC	10620
55	CCGTTAAAAAT	TTTAATTAAT	TATTATATAA	TAAGAGAACT	TTTCAAACAA	TACAGTTGTT	10680

	TTACTGCAAT TATTTTTC	AAATATCAAC GTTAATATAA	CTTCTATTAA GAAATACTCA	10800
	CAATTCTGCCC TGCAATGCAA	ATCTCGTCAC ATATAAATAT	TTTTAATTAT TTTAAAAAAT	10860
5	GATGCACTAA ATTAGCAACG	AGCTTAGCAG TTCTATTGTC	AGCGTCATAT GTTGGATTCA	10920
	TCTCAGCAAT ACTAACTGAA	GACACCTTAT CACTTGGAAT	AATACGTTTT GCTAATTCAA	10980
10	GAACAGTATG TGGATACAAA	CCTAACACTG CCGGCGCACT	TACCCCAGGC GCAAACGCAC	11040
	TATCAATGAC ATCCATACAA	ATCGTAAACA TAATGACATC	ATGTTTCATGT ACAAACGTT	11100
	CAATCATATC TTTAATTGTT	GGTGATACGT GACTCAATAA	TTCATCTGCA AAGACATAAT	11160
15	CAATCTTTTT CTCTTTAGCA	TAATCAAATA AACTTTGCGT	ATTACCACCT TGAGCAATAC	11220
	CAAGCACTAA ATAATCTGTG	TTTTCATCTT CTTCTAAAAT	TTGTCTAAAG CTCGTTCCAG	11280
	ATGTAGATTG TTGTTTCAGCA	CGTGTATCAA AATGCGCATC	AATATTTATC ACACCAATAG	11340
20	ATTGTGTTGG ATAGACTTTA	CGTGTGCTA AATATTGAGC	ATACGCAATA TCATGTCCAC	11400
	CACCTAATAA AAATGTTTTGT	CTATGATTAG CAATTGACTT	CGCTGCAAGC ATAGCAAATT	11460
25	CTTTTTGAGT ATCAATTAAT	TCCTCATGAT CATGATAAAC	ATTTCCGTAA TCGACTAAAG	11520
	TTACATTGA TTCAAATCCG	GCAAACCTGC AAATGCTTGT	TTAATCGCAT CTGGTCCTTC	11580
	TTTTGCACCA ATGCGCCCCT	TGTTTAAAGC AACACCTTTG	TCAACAGCAT AGCCTAATAT	11640
30	ACCGACCCCT GATGGCATA	C TACTCTTTTC CAGCTTAGAC	AAATCTTCAA ATGTTACTGT	11700
	TTGAAAATGT CTAAATTTTT	TCGGGTCTGT TTCACTATCT	AACCTTCCAG TCCATAAATT	11760
	TGGTTCACCT TGCTTGTA	CA CAGCATTTCC CCTCTTATT	TATGTGGCTT ATTAACAATT	11820
35	AAAGTATAAC GTATAGGAAA	TTTTGAATTC AATTCATAGT	TAAATCCGTA TCTTAAAAAT	11880
	ACTTATCTAC ATTACTTTTA	CCCCTATTTT CTATGTAATA	ACGAATACTT AGCTGATTTA	11940
40	TGTTAATAAA ATACGTCAAG	ACTATTACAT TTTCATTAAT	ATTGACATAG ACAATTTATC	12000
	TCTCGGCTTG TAATATGTAT	AATTGTTACT AAAAGATATT	TTGCTTGTTA CCTAATGGAG	12060
	GTTACATATA ATGAAGAACA	ATAAAATTTT TGGTTTTCAA	TGGGCAATGA CGATTTTCGT	12120
45	CTTCTTTGTC ATTACAATGG	CGTTATCCAT TATGCTCAGA	GATTTCAGT CTATAATTGG	12180
	TGTCAAACAC TTTATATTTG	AAGTTACAGA TCTAGACCA	TTAATTGCTG CAATCATTG	12240
	TATACTCGTT TTCAAATATA	AAAAGGTCCA ACTTGCAGGT	TTAAAATTCT CAATCAGCCT	12300
50	GAAAGTAATT GAACGTCTAT	TGCTAGCTTT AATTTTACCT	TTAATTATTC TAATTATTGG	12360
	TATGTACAGC TTTAATACAT	TTGCAGATAG CTTTATTTTA	TTACAATCAA CAGGCTTATC	12420
55	AGTACCTATT ACACACATTC	TGATTGGACA TATTCTGATG	GCGTTCGTAG TAGAATTCGG	12480

TGTGTTGGT TTGATGTATT CAGTTTTCTC AGCAAATACA ACTTATGGTA CAGAATTTGC 12600  
 TGCTTATAAC TTCCTTTATA CATTCTCATT CTCTATGATT CTTGGTGAAT TAATTAGAGC 12660  
 5 GACTAAAGGA CGTACAATTT ATATTGCAAC GACATTCCAT GCTTCAATGA CATTCCGACT 12720  
 TATTTTCTTG TTTAGCGAAG AAATCGGCGA TCTATTTTCA ATCAAAGTCA TCGCCATTTT 12780  
 AACAGCAATC GTTGCAAGTAG GATACATTGG TTTAAGCTTA ATTATCCGAG GTATTGCATA 12840  
 10 TTTAACAACA AGACGAAACC TTGAAGAACT TGAGCCTAAT AATTATTTAG ACCATGTCAA 12900  
 TGACGATGAA GAACTAATC ATACTGAGGC TGAAAAATCT TCTTCAAATA TTAAAGATGC 12960  
 15 TGAAAAAACA GGTGTAGCTA CTGCATCAAC GGTTGGTGTG GCTAAAAATG ATACTGAAAA 13020  
 TACAGTGGCT GACGAACCAA GCATTTCATGA AGGTACTGAA AAAACAGAAC CTCAACATCA 13080  
 CATAGGTAAT CAACTGAAT CTAATCATGA TGAAGATCAT GACATCACTT CGGAGTCAGT 13140  
 20 AGAATCAGCm GaATCAGTTA AACAAGCACC ACmAAGTGAC gATTTaACAA ACGATTCAAA 13200  
 TGAAGATGAA ATAGAGCAAT CATTAnAAGA ACCTGCGACT TATAAAGAAG ACAGACGTnC 13260  
 ATCAGTTGTA ATTGATGCAG AAAAACATAT CGAAAAAGCT GAAGAnCAAT CTTCAGATAA 13320  
 25 A 13321

## (2) INFORMATION FOR SEQ ID NO: 5:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8549 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ATGTGTTGTA AACTTTTATG TTGAAAAGC TACTTATCTC AATGAAAACA AGTAGCATTT 60  
 40 AATAAATTAA TTAGTATACA GCTAGTTTTT CTAATTGTTT TTTAACTTGA ATTAAGTTTG 120  
 ACCGTATTAG AGAGGCAGAT TGATCCATCG TTTGAATTGC TTGTCCTTCA TTTTCGTTCA 180  
 45 AGCCATTACA AACAACTTCA AACTGTTGTG CCATTTGATC AAGACGCGCA TGAGCTTGTG 240  
 TGTTTAAAAT AAACATATCG TCATAATGTG ATGGCGAATA GATAATTCGT CGTTGTATAC 300  
 AAACGTATAA AAACCTTGTC ATATCAACGG TTTTGGCATT TTAAACCTC TGTGTTTTCC 360  
 50 ACGCATGTTT GCCCTTATTT AAATAATTTG CCCTTTTTTC GCCCCGAAAA AAAACACAA 420  
 AAAAATAACC AACTCCTAA ATTAATAGGT GGTGTGGTTT TGTTGATTGT AGGGGTATAA 480  
 AAATAACCGC ATTATTAAAG ATACGGTTAC TCTGTTATCT GTAAATATAA TAGTAGTTTA 540  
 55



	AAACAGGACT CCACATAAAA ATCAACTCCT TTATATACCA TAATGATACT ATATTTTCTA	660
	GTTTATTTCA ATTTTTCAGT TTTTAAAAAT GAGTTTCTGT TTTTATTTAT ACGCTTTTCT	720
5	GTTTTCTTTT TAAATTTTAT CTTTTTGTTA TTCCATTCTAT TGTAAAATTC TATTAAATTA	780
	ACATAAAAT TTTTCATGCCC TATTTTATTT GTTGATGAGA TATCAATGTA AAGACTCAAT	840
10	ATTGTTTTTA AATAGATTTG ATGCAACGAC TGATAAACCG TATTACTATC TGCTATGTTA	900
	TTGGTAAAAT GCATAGAAAA ATATTCTAAT TTATTCATGC AATATATATG GGTTTCATTA	960
	TACTTCTTAA TGAGTGTATT TATACCTTGC AATACGTCAT TACTTTTAAT AACAATTTCT	1020
15	TTTTCACCTG TCGAAAAAGT CCACTGTTTA TCTCCTATAT TTTCTTTAAT TGTTTTCTTG	1080
	TTGTCAAAT CTAAAATTAT AGCCCGTAAA CACTCTTCTT TATAATTCTC GTTCTTGAAA	1140
	GTACGAAGCA AAATTTTTAT AAATTCGGTA TTGGTGACTT TTTTATAAGT GTGATATTTT	1200
20	GCAATCTCTT TATCAGTAAA GACTGTTCTT AGTTCGTGAT TATCAAAACT TAAATTCATC	1260
	TTATTCTCTA ATTCATTAAT TTTATCTTGC AAACCAACAT TTTCTAAAAT TTTCTTGTTT	1320
25	ATCTCCCCTA TATCAAAACT CCTTTTCGAA ATTAATTTTG AAAACTCGTC TGCCATTTC	1380
	ACAGCCTTTT CTTTCCTTTT ATACCTTTTG TTAAATTTAT GAACCACCGT TGCAGCATAA	1440
	TACGATATCC CACCAGATAA AATAGATGAT ATTATCGGTA TGTATATATC ACCTTTCATA	1500
30	TTTCCACCTC TTTTAACACA ATTAAGTATT ATGATACACA ACTTGCGCAA AAAGATGTAG	1560
	ACAGAACATA ATGGCGAACA AAAACAACCA CCCAGTAACT AGTATGGGTG GCGTAGACTA	1620
	TAACAACTCT ATGTTATCAA GATATATGTA TCGAGTGATG GCAAGGAAGA AGTCTCCTGC	1680
35	GGGACCAACA GTCAGATATA TGGCCTCTGC CGGGCTATAT AGTTCACTCC TACTATATAA	1740
	AAGTAAGTAT AACATAAAAA GCACCCCGTA AACTGTTATA CGGGAATGCT AAAGTCATAT	1800
40	ATACTACGGG GAGTAGTATG AAAACTATGC TCTCTATCGT AAGAAAAAAC ACCCAGTGAC	1860
	ATGCTTGGGT GAACAAGGAT AGATGTAAAT AGTTGATGCA TGTGTACACA TCATAACAAA	1920
	AAACTAGCCC GAAGCTAGCT ATAACATAAA AAAATAGGCA AGTACCGAAG TACCTGCCAG	1980
45	TTACGCACAT TTAAATCTTG AGAGTAATGT TAAAAAGTGT ATAGGAATAT TAACATCCAT	2040
	CCAAATAGTT ATTTAATAAC TGTAAGATTC CCTATAATTA ATGTAGCAAA ATTTTATTTC	2100
	TAAGTAAATA CTAAATCGTG CTAAACTTAC CAAACTACT TATTCTATTA CCTGCCTTGT	2160
50	CTACCTCTCC TGTGCTATA TAACGACGTT GTCCACTATT AGCAATATAA GTAATCCATC	2220
	TATAGCCATT GATGCAATAT GCGCCGTCAT ATTTAATTGT TGC GTTATTA GGTAATACAC	2280
55	CTGTAATTCT TGAATTAGTT GAATAGCCGT CCCTTACGTT ATTACCTTTA ACATTGGCAA	2340

	CTGGCACTGG	TGGATTTTTT	TGGTTTTTAG	CTGATGTTTT	AACATTACCA	GCTACCAAAC	2460
	CACCTATAGG	CTTACCATGA	ATCGCACCGG	CTATTAATTT	AGAATACAAG	TCATAGTTTT	2520
5	TCTTAATCCA	ATCCATATCA	TTTTTATTAG	TAATAAAACC	TAATTCAGAT	AAACGATAGT	2580
	TTATATTTAT	TTCTGCTGAT	ACATTAACGT	TTAGTAAATC	ATTACGAGGT	GTTACACCTC	2640
10	TTATTTGTCC	TAAGTTATTT	TTAATAACAT	CTTGATACT	TTTATCAATA	GTATCTGCAT	2700
	TGAATTGACT	TGAAATAATA	ACATGCCAC	CACTTGCACT	TTCTCCTGCT	GCGTCTAAAT	2760
	GAATCTCTAG	AACAATGTCA	TACCCATGTG	ATTTAACCCA	ATATAAGCCA	TAATCTTTAT	2820
15	TATTTCTTAC	ATTAACACCG	TAAGCAGTAT	CTTGATACAT	ATCTTGATGAT	TGACTTGAGC	2880
	CACCATATAA	TGCAACTTCG	TGACCTGCAT	GTCTTAAATA	CTTAGCGATA	TTTGGTGTTA	2940
	TATATTTACG	GATAAAATCA	CGTTCATTG	TTCCGTTTCC	GACTGCTCCA	GGATCGTTAT	3000
20	AACCATGACC	GGCTACAAGC	ATAATTTTTT	TAGGTTTAAT	TACTGCTTGC	TTTTTGGCAG	3060
	TTGCTTGCTT	AATAACGCTT	TTAGCTTTAT	CTCCAACACT	TACTTTATCT	GGGAAATTTA	3120
25	ATCTAATAAA	ATACATTGGG	TCATCGTAAT	AATGAACATG	TCTTGTAACG	GTTTCGGGAC	3180
	CCCAACCAGG	TTGCGCAACG	CCATTTGTCC	AACCTTTACC	ATTCCAATTT	TGGCCAAACG	3240
	ATGTGAAAGT	GTTTAGATTA	GCGCTCTCAA	CAATTTCAAC	ATGTCCaGct	CCGCCACCAT	3300
30	ACTTTGACGG	GAAAACGACA	ATGTCCAAC	TTTGCGGTAA	AAAGCTATCA	TAGTTTTTAA	3360
	TTATTTGCCC	GTATTTTTC	ATCCTTGCTT	TATTATCAAA	TGGAATATTA	TAAGCGTATA	3420
	AACCTTGTA	CcTTTCGCCT	GTTGCTATCA	TAAAAACAT	ATTTGCGTAA	TCGTAACACT	3480
35	GAAATCCATA	AAACAAATCA	GGATTGAACT	GCTTCCCTAA	TGAATTATCA	AACCATTTTT	3540
	CTGCTTGGTT	TTTTGTTATC	AACATTGGTC	AACACCTACC	CTAAATCATT	TGTGTCGTTT	3600
40	ATAFTCGTAG	GTGTCATTAC	TTCTTTAATT	GGCGCTTGCC	CTGTTGCTTT	TCTATACTTG	3660
	TTTTCAGCTT	TATATTTCTT	TAGCTTTTGA	TTTGCCCAT	TACCTTCTTG	AGATGTTGGA	3720
	TTATCTTTAT	ATGTAGTATA	TAAAGCAACA	ACTGTTAAGA	TAATCGATGA	AACACTTTCT	3780
45	TCATCTACTG	GTATCGGACT	TATACCTTTA	TTGCTAAAA	ACTGATTGAC	TAATGCTAAG	3840
	ATCAATACGA	TGTATCTTGT	TATTACTTTT	GCATCCATTT	GTTTGCTCCT	TTTATCCAAA	3900
	ATAAAAAGCC	AGTGCCGAAG	CACTGACTCT	TAACTATTAC	TTACACTTAC	TAAACCAGAA	3960
50	ACACGACCAA	AAGCTATATC	CTAAAATTCC	CTTAAGCATG	GTAATCACCT	CCTTTAAATG	4020
	CCAAAAATAG	TTTTTAACAA	GGCTATAACA	AATGTACTTA	GAATCGTCCC	TATTAATCCT	4080
55	AGAATCCACA	TCTTGATGTC	TCTAATATTT	TTAGCATTTT	TCTCTTTATT	TTTTTCATCT	4140

EP 0 786 519 A2

	TGCGTTCTCA	GACTGTCTTC	TATTCTGTCG	AATTTTTC	ACATAGTCTT	ATCATTCTCT	4260
	TCTAATCGCG	TTAAACGCCA	ATCTTGTTTC	TGTCGTTTGG	TAAATCCAAA	CATTACACCA	4320
5	CCCACTTTAT	TCAAATTAAA	AAGCCATAAG	ATTATAACCT	ATGACTCTAG	ATTTTCTGGA	4380
	TACTTTTCTC	CTGTAATAAT	TGCATATTCC	TCTTTATCTA	TAACTTCCAT	ATCTACATAC	4440
10	CACGCTATAT	CTTCTTTACT	ATATTCTTTC	AATTGATACC	ATGTTTAAAT	ATCTTCGAAT	4500
	GTTGGTGAAA	TTAATTTAAG	CATTTTCAGT	CTCTCCTTTA	ACCTCTCTTA	ATTTTTTATT	4560
	AAGTGTCA	AGTTGTTTTG	CCATTAGTGC	ATTTTGCTTA	TTAACTTGCA	TCGATAACTT	4620
15	TGTACTTTGA	ACAACTTGTT	TCTGCATACT	AGCAACCATT	TTTCGTAAGA	TGTCATCAGA	4680
	AGCGACTGTG	TTTTGTCTCT	CACTGTCAAT	CTGTTGATGC	AAGTCATCTT	TTTCTTCTGA	4740
	ATAATCTTCG	TTAAAACTA	TTTCCCCATT	TGAATATTTA	AAGGCTTTAG	GTCTAAAAAC	4800
20	TTGAGAGAAA	TTTTCTGGTA	AATTTTCAAT	ATCAATACCT	TCTTCAAAGC	CACCAATGAT	4860
	AGCGTATGAA	ATTATCTCAT	TACGCTTGTT	AACTAATATT	TGCATTATTT	TCTCACTCCT	4920
25	ATAATTTTGT	TAATTGTCCC	TCTATTTGCG	TTTCGCACCAG	AGCCTCTTTG	ACTTCCTAAG	4980
	TCGAAATAGA	CATCGTTTGA	TATAGTTAAA	GATGTACGAC	TAGATTTAGT	TAATCCAAAC	5040
	TCATAAACAC	CTCCACCATT	TCCATCACCA	TCTGGAAGAT	TTGAGGGATT	CAATGAAATC	5100
30	TTTCCTCCTC	CAAAGGACT	GCCAACTCT	GTAAAGTCAC	CACCTGGAAA	AGTCCCATAA	5160
	AAAATTAATA	AAATAAATTG	GTCTAAACTC	TCATTTAAGT	ACAATGTAGA	GCCACACCA	5220
	TTTGCTGTTT	CATCAAAAAT	AACCGAATAC	CTTTTATTAA	ACTTGTATC	TGCGTATAAT	5280
35	TTAGCGTTAC	TTTCGGCCAT	ATTAGCTTTT	GATTGGGCAC	TTTGAACAGT	TTCAAAGGT	5340
	GTATTGTAAT	CATTAATAGC	TAATTCTGAC	CACTCAGACC	ATGAACCCGC	TTCTTTTCTT	5400
	TTAACAAATA	CTTTATTTGT	ACCGTTCCGT	CGATAAGTCA	TACGCTTGTA	ATCTGAAGTT	5460
40	ACTACTAAAT	ATTCGACAGT	ACCGTTAGTA	CTAACACCTC	TTGGATAATT	TATAGCTTGC	5520
	GAAACATAAA	TAAATTGGGT	TGAATCACCT	ATTCTTTGTT	CTGGATTATT	AAAATCAAAT	5580
45	CCAGTAATCT	GCATTATCTT	ACCATCATCT	TTAGTAATCT	TAGCTTTTGT	CCAATTTGAA	5640
	GTAGAACCAC	TTGTGACTAA	ACCACCACTA	TTCACTGACT	GCTTGAAGGC	TTTATGTTTC	5700
	TCATCCATAT	ATCGCTTTTG	CTCATCGAAT	GTTCTTGAAT	ATGCTTGC	TTTATTTTCC	5760
50	AAATCAGATA	TATGGCTATT	AGCAAGTTGC	TTTAATTCAT	CTATACTTGA	AGATTTTGCT	5820
	ATTTGAATAT	CTGATAGACC	TTTTCTTTTA	GCTTTTTC	TCAGACTCGC	ATAATCTTCA	5880
55	CCATTTTTTA	TAGCCTCGTC	CATTGCTTTC	GCACGATCCA	TAATAGTTTT	TTCTAATTC	5940

	TCAACGTTAA ATGTGATAGT TCTCTCGACA ACTACCACGT CTGAATTACC TAATTCTGCA	6060
	ACCGAAACTT GAGCTTGATA ACTTCCATCT CGTTTAAATTA CATCATTAGG TAATTGAAAT	6120
5	TTTAAATAC CTTTAAATGG ATCTAATATT TCTAGTGGAG CAACTACCAT GACTCCTTTA	6180
	CCTCGAATCG CTATTCGTGC KTTGATATTT tCTTCACTCA ATAATAACGG TTGATTATTT	6240
10	TTAGTGATAT TAAAAAGAAG AACAGAAGAA TCACTCTCTC CTGTTCTAAA AGTTATATCT	6300
	AGATTGAAA TATTTCCATA ATGCGCTGTG TTTTCTAAAT TTATAGCTAC AGATTTCTCT	6360
	AAATTACTCA TTAACCTATA ATTCTCCCTT CGTGTAAGT CCATGGCCCT GAACTTGTTT	6420
15	TACTATCATA ATTTTTCAAT AGTATCTCAG CAGATGCTGT AACACTATTA CGAACTAGCC	6480
	TATGAACAAA GCCACCTGTG TTTGAAGCTT CTACATATAA GTTCCAACCA GCTACCCCTT	6540
	TACGTTCACT TGGAAAATCT GTAAAACGTT TTGTATCATC CGTAGTTAAA TAAAACGACA	6600
20	TGCCTACTAT GTTAATATCT GACATTTTTG TGATGAATGA AGGTACTCTC TCCCATTTAC	6660
	CACTATTTTT AGGCACATAA TTCCAGTCCG AAATGTCTCC AGTTCTTCCA GAAAGCACCC	6720
25	TTTCAAAGT CATCATATTC CTTGCATAAC TATTACGCGT CAATATCTGA ATTACATCAC	6780
	CGCCAGTTTG TGGTGGCTTA ACTTCCAAGA ACCAACCTGC ATCACGCCAT TCTCTTGGTA	6840
	ATGGGAAATC ATCGATTGA ACTGTATGAT CAGTGTATAA ATAGTAAAGA CCTGGCTCTG	6900
30	TTAACATCCC AAGATTCTTA AGTTTATCAG GCCTCATTGG TAAAGGTTTA ACTCTACCAC	6960
	CTGTGTCACT CaTGATAAAA GGAACGCCTC TTGAGTGAAG TATTTCTAAA ATACCTCTTT	7020
	GCCCAATCAT GAAAATACGA TGTGTTCTAT TTCCaTCACC ACCGACAGTA ACACCTAGCA	7080
35	TCAAAGCTTT TTTACCACTA TCTTTGTCAT AGTATATTTG CAAACCTTtC TgCTTCCGCA	7140
	AATTCGCCAG GAAATGAATC tAgTGTTCCA CCATAGTCAG CATTAACTG ATACGCTTCT	7200
40	TCTCtGTTT CTAAATCGAA AGCCGTTAAA TAGTTTCTAT TATTTGGATT ACTGTCTCCT	7260
	GTATACCAAT ACAAGTATTT TTCATCAAAA GTCACACCCT GCATTGGTTG GGTTCGTTT	7320
	GTTAGTCTCA TAGGGATACT GATTTTATGC AAAACTTTAT CAATATTTTT ATCAACATCG	7380
45	TCTAAACTTC TTATCTCTAT ATAAntCATT GAGTTTTCAA GTTCCCACTG ACTTCTAGGT	7440
	CTCTCaATTC TGTATAGAAT TTTATTTTCT TTTTCATTTA TGACAGGGGT GATGTAGGGT	7500
	TTTTCTGGGT GTCCTGTAAA TACATCTTGC ATACCATACT TGCCATAGCT AATTTCCACA	7560
50	TTAGGCGTAT ACTTGAAACG AACTAATGTA TTCTCATTAT TACCATTTAA GATAAACTA	7620
	TAAATCCATA ACTCATcATC AATATATCTA TAACCGTTAT GTGTACCATG ACCCCCACCT	7680
55	ACAATCAATG AGCTGTCTAT AAATTGACCA TTAGGTCTTA GACGACTTAG CATATAGCCA	7740

ATTACTGCAT TTGTAAGAGG TGCAAGTTCT GTCACAAATA AAAATTCTTG CTTATCAGGT 7860  
 TCAAAACGAT ACTCGATATC AAGAATTTCT TGTGGTCT TATTTAATTC TCTTATAGTT 7920  
 5 TCCTCTTTAT TAATTGAGT TTTGGTTTCC CAATCGTCTA AATGTTCTTT TAATGTGTCA 7980  
 AAGGTTTCGC CGTTTACATT AACTCGAGCT TGAACAATCT CATTAGCACT GTTATTACGT 8040  
 10 GGTGCCACAA CAAGTGC GTT AATTTGACTT TGTAAGATT TGTCTACTGC TGCTTGCAT 8100  
 CTACCATTAT AATAAATTTG CTCAGCGAAG TGTGAATTG TTTTAGCTyT CTGATGCAAC 8160  
 TTAAACTCTG TTGTCAAGCC AAGCGCAAAT TGCTCTATTC TTTGTAAGTT TTGTATTTCC 8220  
 15 TTAGCTCTAT AATCTCGACC TGCTAAAGCT CCCAAATCCT TTATTAAATA CAAATTTTCC 8280  
 ATAATGCACC TTCCTTTCTA ATAAATAGC ACTGTACCAA GTTCCCACT ATCGTCAACT 8340  
 GTTATTTTCC ACAATTTACC GTTTGGGGAT TTCTGTACAA TGCTATTTTG AATAATTgcC 8400  
 20 TGctTCGCCT ATTTTAAAT TATCTAATTT ATTTktATCA TTTACCGAAA TGATACCGTC 8460  
 TTGAGGCAAT CCATCAATAn CACTACTGCC TGCATAAGGT ATCCCATTTA TAGCTTTCCA 8520  
 25 ATGTGTAGCT GGAAAGTACT GTTTATCGT 8549

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3601 base pairs  
 30 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AGGCGTGTAG TGAATTACGG nTAGGAACT ATGTATCCGA ATGATTTATT GAGACCAAAA 60  
 AGGCATTAAA GTCCATTGAA ATATChGGTA GCGmGTTGGT ACgTGGACGT GGGGGCCCTA 120  
 40 GATGTATGAG TCAACCATTA TTCAGAGAGG ACATTTAACG TAATAAATTA TAGAmACGAG 180  
 GGTGAAAATA ATGACAGAAA TTCAAAAACC GTATGATTTA AAAGGCAGAT CATTATTAAA 240  
 45 AGAAAGTGAT TTTACCAAAG CAGAATTCGA AGGACTTATT GATTTTGCAA TTACATTAAA 300  
 AGAGTATAAG AAAAACGGTA TTAAGCATCA CTACTTATCT GGAAAAATA TTGCACTACT 360  
 ATTCGAAAAG AATTCGACGA GAACGCGTGC TGCCTTTACA GTTGCCTCTA TTGATTTAGG 420  
 50 TGCGCATCCA GAATTTTTAG GAAAAATGA TATTCAATTA GGCAAAAAG AATCTGTAGA 480  
 GGATACTGCG AAAGTATTAG GTAGAATGTT CGATGGTATT GAATTCCGTG GTTTTTCACA 540  
 55 ACAAGCTGTT GAAGATTTAG CGAAGTTCTC TGGTGTACCG GTGTGGAATG GATTAACAGA 600

	TCTAGAAGGA ATAAACTTAA CTTACGTTGG AGATGGACGT AATAATATTG CGCATTTCATT	720
	AATGGTAGCA GGTGCTATGT TAGGTGTAA TGTAAGAATT TGTACACCTA AATCATTAAA	780
5	TCCAAAAGAG GCATATGTTG ATATTGcAAA rGAAAAaGCG AGTCAaTATG GTGGTyCAGT	840
	CATGATTACG GATAATATTG CAGArcCAGT TGAAaTwCm GATGCTATAT ATmCAGATGT	900
10	TTGGGTATCG ATGGGTGAAG AAAGTGAATT TGAACAcGTA TTAATTTATT AAAAGACTAT	960
	CAAGTGAATC AACAGATGTT TGATTTAACA GGTAAAGATT CAACGATATT CTTACATTGT	1020
	TTACCAGCAT TCCATGATAC AAATACACTT TATGGACAAG AAATTTATGA AAAATATGGA	1080
15	TTAGCTGAAA TGGAAGTTAC AGACCAAATC TTTAGAAGTG AACATTCAAA AGTGTTTGAT	1140
	CAAGCTGAAA ATAGAATGCA TACAATTAAG GCAGTAATGG CAGCAACATT GGGGAGTTAA	1200
	TCACTAAATG GAACGATATG AATATGATGT GTCTGATGAT ATAAGTGTCa TGTACAGACA	1260
20	CCTCATATTG GTATTAAAGG AGAAATGAAT ATGAACGAAT CAGGAGATAA CAAACTCAGT	1320
	AAATCTTCTT TAATTGGACT AGTTATAGGA TCCATGATTG GTGGCGGTGC GTTCAATATA	1380
25	ATGTCTGATA TGGGCGGTAA AGCCGGTGGA TTAGCCATTA TTATTGGTTG GATTATTACA	1440
	GCTATAGGAA TGATTTTCATT AGCGTTCGTA TTTCAAATTA TAACCAATGA ACGGCCGGAG	1500
	CTAGACGGTG GTATTTATAG TTATGmTCAA GCAGGATTTG GCGATTTTGT AGGATTTATC	1560
30	AGTGmTTGGG GATATTGGTT CTCAGCGTTT TTAGGCAATG TTGCCTATGC AACACTATTG	1620
	ATGTCAGCAG TAGGTAACCT TTTCCCGATT TTTAAAGGAG GCAACACATT ACCAAGTGTT	1680
	ATTGTCGCCT CGTTACTACT CTGGGGTGTC CATTTCTTGA TTTTAAAAGG CGTTGAAACA	1740
35	GCAGCATTTA TCAATAGTAT TGTTACTGTT GCAAAGTTAA TACCGATTTT ACTTGTAATC	1800
	ATATGCATGA TAATTGCATT CAATTTTGAC ACTTTTAAAA CAGGCTTTTT CAGTATGACG	1860
40	TCAGAGGGTG TATTGCCATT TAGTTGGGCG AGCACAATGA GCCaaGTtAA AAGTACGrTG	1920
	CTAGTGACAG TTTGGGTGTT TATCGGTATC GAAGGTGCAG TAATTTTTTC TAGTAGAGCT	1980
	nAAAATGAGA AAGATGTAGG TAGTGCCACG GTTATAGGAC TTATATCAGT TTTAATTATC	2040
45	TATyTCTTAT TAACTGTATT AGCTCAAGGC GTGATTTTGC AAAATCATAT TTCGCAATTA	2100
	GATTCGCCAA GTATGGCACA GGTGCTTGCA ACTATTGTAG GTGGTTGGGG ATCTACACTT	2160
	GTAAATATTG GTTTAATTAT TTCGGTACTA GGTGCATGGT TAGGATGGAC ACTGCTTGCT	2220
50	GGTGAATTAC CTTTCATTGT TGCAAAAGAT GGATTATTTT CAAAATGGTT TGCTAAAGAA	2280
	AATAAAAATG GAGCACCTGT AAATGCACTG CTTATTACCA ATATATTAGT ACAATTATTT	2340
55	TTAATAAGTA TGCTATTTAC ACAGAGTGCG TATCAATTTG CATTTTCACT AGCATCAAGT	2400

CGACAGCAAG CAACTACTAA ACAATGGACG ATTGGTATCA TAGCCTCAAT TTATGCTATA 2520  
 TGGCTTATAT ATGCAGCAGG TATCAATTAC TTATTATTGA CGATGTTACT TTATATTCCA 2580  
 5 GCTCTTCTTG TTTATACaAT CGkTCmAAAG rATwATCAGa CACGTTTGAT TAAATCAGrC 2640  
 TATATTCTtT TTATGATTAT tATCGTACTT GCAGTTATCG GGTTAATTAA GTTATTGATG 2700  
 10 GGAACGATAA ATGTTTTTTTA AAAGGAGCGA CAAAAATATG AAAGAGAAAA TTGTCATTGC 2760  
 ATTAGGCGGT AATGCGATAC AGACAACAGA AGCAACAGCT GAAGCACAAC AAACAGCTAT 2820  
 TAGATGTGCG ATGCAAAACC TTAAACCTTT ATTTGATTCA CCAGCGCGTA TTGTCATTTT 2880  
 15 ACATGGTAAT GGTCCACAAA TTGGAAGTTT ATTAATCCAA CAAGCTAAAT CGAACAGTGA 2940  
 CACAACGCCG GCAATGCCAT TGGATACTTG TGGTGCAATG TCACAGGGTA TGATAGGCTA 3000  
 TTGGTTGGAA ACTGAAATCA ATCGCATTTT AACTGAAATG AATAGTGATA GAACTGTAGG 3060  
 20 CACAATCGTT ACACGTGTGG AAGTAGATAA AGATGATCCA CGATTTGATa ACCCAACTAA 3120  
 AcCAaTTGGT CCTTTTTTATA CGAAAGAAGA AGTTGAAGAA TTACAAAAAG AACAGCCAGA 3180  
 25 CTCAGTCTTT aAAGAAGATG CAGGACGTGG TTATAGAAAA GTAGTTGcGT CACCACTACC 3240  
 TCaATCTATA CTAGAACACC AGTTAATTCG AACTTTAGCA GACGGTAAAA ATATTGTCAT 3300  
 TGCATGCGGT GGTGGCGGTA TTCCAGTTAT AAAAAAGAA AATACCTATG AAGGTGTTGA 3360  
 30 AGCGGTTATA GATAAAGATT TTGCTAGTGA GAAATTAGCA ACGCTGATTG AAGCAGATAC 3420  
 CTTAATGATT CTTACGAATG TAGAAAATGT ATTTATTAACT TTTAATGAAC CTAATCAACA 3480  
 ACAAATCGAT GATATTGATG TAGCAACACT GAAAAAatAC GCGGCACAAG GTAAGTTTGT 3540  
 35 GGAAGGATCG tGTTGCCAAA AATAGAAGCT GCGtACgtTT GTTGAAAGtG GGGaAACCAA 3600  
 A 3601

40 (2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 573 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

50 CGACACTATT AAATGAATTA GAGCACAATC TAACAAATCA AATTCATTTT TCAAAAGATG 60  
 AACGACTCAC ACATATCGCT TTAAAGTTAT TCGAAACAAC CGATCCTGTT TCAACAAAGC 120  
 55 AACTTGCGCA AGATGTTAAT GTTTCGCGTC GGACAATTGC AGATGATATT AAAATGATTC 180

TTATTGGTGA GGAAGATCAT TATCGTAAAG CGTATGCACA CTTTATACAT CAATATATGA 300  
 AACAAAGCTGC ACCTTTTATA GAGGCGGATA TCTTTAATTC AGAATCAATC GCATTGGTTC 360  
 5 GCCGTGCCAT TATTAAGACA TTAAATAGTG AAAATTATCA TTTAGTTCAG TCGGCTATCG 420  
 ATGGCTTAAT CTATCATATA CTCATTGCCA TTCAGCGTTT AAATGAAAAT TTTTCGTTTCG 480  
 ATATACCTAT CAATGAAATT GATAAATGGC GACATACTAA TCAGTATGCh ATTGCTTCAA 540  
 10 AAATGATAGA AAACCTAGAA CGCAGTGTAA TGT 573

## (2) INFORMATION FOR SEQ ID NO: 8:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1221 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TTGATATTTA TAACGTTATA TTTTAATAGT TCACCTGGAT TATTAAATAA ATAGTCCGCC 60  
 25 AAATTTTCTT TTTCTTTATC AATCTGaTkG TAATTAACaC TTTCGaCTTC TGTagGAATT 120  
 CTAATGTCAA CAGAAGCATT GATATAAGCT TGATGTTGCA TGCAATCACA CTCCTAATCC 180  
 30 TTCATmTmAA ACGGAGAAGT AAACCCGTCA CTATTCAAAT TCAATCCTTT TGCCCAATCA 240  
 ACAGGCTTAT TCATGATAGT TTCGATTTCC TTAAGTCCAT TTGAACCTCT AGGTATTTCT 300  
 ACAATTACTT CATCATGGAC ATGGCCAAC TTTTAAAAC CTAATGCTTC AAGCCTTGCT 360  
 35 ATAGAAATCG CAAGTAAATC CCTTGCAGTT GCTTGAACAA TATTCTCGAC TAACTTCCCA 420  
 CCATACGTTT TTAACCTTGA CCATTTACGG TTAAGATCTA ACCCCATAAA TTCAACAACT 480  
 TGA<sup>5</sup>CTACCCC AACTATTTTC ACCAACTAAA GCTTTTGGAT AAGCTAAAGC TCTTCCACTA 540  
 40 GGCAGTTCAA TCATTAGAAA ACCTTTTTTC ATATAAAATC TAAGTCCATG TGTATGATGC 600  
 GTCTTTCTGGG ATTTTACAGT ATTAATTGCA GCCTCTTGGC AAGCCTTCCA AAAATTAACT 660  
 ATGTTAGGAT TTGCGTTACG CCAACTATCA ACTAAACCTT GTAACCTGTT TTCTTCAATG 720  
 45 CCCATTTCCA ATGCACCCAT TGCTTTTAAA GCTCCAGCGC CACCTTGATA GCCTAAAGCT 780  
 AATTCGGACA CTTTTCCTTT TTGTCTGAGA GGGTCGCCTT TAGTTATGCT TTCTACCGGT 840  
 50 ACATTAAACA TTTGAGAAGC CGATGCTTCA TATATCTTTC CGTGTGTGTT GAATACATCT 900  
 AAACGCCATT GTTCTTTTGC ATACCATGCT ATGACTCTTG CCTCTATTGC AGAAAAATCA 960  
 CTTACTGCTA GTTCATTACC TTCTTCAGCA GTAAATGTG TCCTAACTAA TTGACTTAAT 1020

55



AGATCTCTTG CTATTTCTAA TTCAGTATCT GAAATATAAT GCTTTGTTAA ATTCTGAAGT 1140  
 TGTACACCTC TACCTGCCCC TCTTCCAGTA CCGGCACCGT AAAATTGAAA CAGACCTCTT 1200  
 5 ACCCGTTCAT CACTGCACAT C 1221

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1090 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTTTGTTTGG TATGAGGTAG CAATGACGAC GTGTCATTGG TGGAGATTGT AAAAATACAT 60  
 20 AATAAAAAGA AGCGGCAATG TATACCGCTC CTTTTTTTATA CTACATACCG ATTTTCAACC 120  
 ATCTCTTTCT ACTTAGTAAT AAGACAATAG TATTAECTAT AAATAGAAGA ACGAAGAATG 180  
 ATACTATATT TATAATTTCA GTAGGACACA TAAATGTTGA CTCGTTATTC AATATTTTTT 240  
 25 CTACGGCACG ATACATCGTA TTGCTCGCCT CAAATGGAGC AACGATACCA AATATATTTT 300  
 TATTAATGGC AACTAAGATG ACTGAACCAA TCCAATATAC AATGCTGATA CCTAAGCTGA 360  
 30 TTAAATGTT AGGTGAAACC ATACTAATCG TTCCAACAAC TAAGATATAT TGTAAGATAA 420  
 CGAGTGAAAA TAAGATTATT AATAGTAAGT AATGTGAGAA ATCCGAATAT ATAATTGAAA 480  
 TAATAGTGAT ACTTAGAATT ATGAACACTA AACATTCAAA AAATAACACT GCTACCTTTT 540  
 35 TATAGAAGAA GGTAAAGATA TTATCGCCAA TCAATTTATA AAACAGGATA TTTTATTTCG 600  
 AATACTCTTT ATTAATAAAA TATGCAATAA CAAATGAAAA TAGTAAGAAC CCTAATTGCG 660  
 TTGCAACAGT ATATGAACTG AAGAAAAACT GGCTATAGCT TAACTTTTA ACTTTGTCTA 720  
 40 TACCTATTGG TAAAAAATAC CCAAGTAAGA AAAGGAATGT GAATAGCACA ACAAGCGTGT 780  
 AAATAATTTT ATTGGAATA CTTTTTTTAA ATTCTAATTT CAAAGTGGAC ACCTCAATTA 840  
 TAAATTAATG TAATCATTTA TGACTTCTTC TTTTGATTGG TACTCTTCTA TTTGAAGGTC 900  
 45 TTTAAAAATA AAGTATTTAC CCGGCAAAGC ACTTAAATCG GATAAATTaT GTGTAATATT 960  
 GATAATAGTT TTAGTTTGAT GGCTTTGAAT AAAATCATTT AAAAATTCAT AAATTCATT 1020  
 50 AACTGTTTTT TTGTCTAAAG CGTTTGTAAC TTCATCTAAT ATGATTAAAT CATGATCTTC 1080  
 CAATAAGAAA 1090

(2) INFORMATION FOR SEQ ID NO: 10:

(A) LENGTH: 904 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

10 TTAGGACTAT TTTATCATAT TCATTTAAAT TACGGCTAAA AATTTTAAAA ACGGGGATTA 60  
 ATATATGGAA TTAAGCTATG AAAGTTAATT GATACTTGCA TTTTACGCTG ATTTATATAA 120  
 GAATAACTAT TGTATAGTTT TAAAAACGAA CGTACGTTTG CAGGAGGCCA AATCATTGGC 180  
 15 AATGAATAAA CAAAATAATT ATTCAGATGA TTCAATACAG GTTTTAGAGG GGTTAGAAGC 240  
 AGTTCGTAAA AGACCTGGTA TGTATATTGG ATCAACTGAT AAACGGGGAT TACATCATCT 300  
 AGTATATGAA ATTGTCGATA ACTCCGTCGA TGAAGTATTG AATGGTTACG GTAACGAAAT 360  
 20 AGATGTAACA ATTAATAAAG ATGGTAGTAT TTCTATAGAA GATAATGGAC GTGGTATGCC 420  
 AACAGGTATA CATAAATCAG GTAAACCGAC AGTCGAAGTT ATCTTTACTG TTTTACATGC 480  
 25 AGGAGGTAAA TTTGGACAAG GCGGCTATAA AACTTCAGGT GGTCTTCACG GTGTTGGTGC 540  
 TTCAGTTGTA AATGCATTGA GTGAATGGCT TGAAGTTGAA ATCCATCGAG ATGGTAATAT 600  
 ATATCATCAA AGTTTTAAAA ACGGTGGTTC GCCATCTTCT GGTTTAGTGA AAAAAGGTAA 660  
 30 AACTAAGAAA ACAGGTACCA AAGTAACATT TAAACCTGAT GACACAATTT TTAAAGCATC 720  
 TACATCATTT AATTTTGATG TTTTAAGTGA ACGACTACAA GAGTCTGCGT TCTTATTGAA 780  
 AAATTTAAAA ATAACGCTTA ATGATTTACG CnwGGgTAAA GAGCGTCAAG AGCATTACCA 840  
 35 TTATGAAGAA GGGATCaAAG rGTTgTTAGT atGTCCAaTG ArGGAAAAGA AGTTTTGCCT 900  
 GACG 904

## (2) INFORMATION FOR SEQ ID NO: 11:

40

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11271 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

50

GATTTCTAAA TCAAGATCTG TTTTACGATA ACCATTCAAA CCTTGACGTT CATCTTCTTC 60  
 AGGTTGATTT TGTGCTGTG TGTCTTTGTT GTCAGAAGTC GCTACTGTTT TTTTATTATC 120  
 55 TGTTCCTTTA GTCATAACAA ACGCCTCCGT TATAAACGC TATATTTAAT GATATGTGAT 180

	TTAATAAGAC GATTCAGCAA GTTTTAAAGT ATTATTTGAC TATGTTGGAT TAGGCATCTA	300
	GTCCTATAAT ATCACTGACA TTGTCAAAAT GATGATCTTT TAAGTAACGT GCGATGCCTT	360
5	TGTTCAATTT CTTAGTTAAA CCTGGGCCTT CAATAACAAG TGATGAATAA ATTTGAATAA	420
	GTGACGCACC GTGACGCATC ATTTTGATTG CATCTTCAGT ACTGAATACG CCGCCTGTAC	480
10	CTATAATTAA AAATTCACCA TTTGTTTGCT GATAAgCATA CTTAATCAAT TTTAAATTAC	540
	GTTCAAATAA TGGACGACCA CTCAAACCGC CTTCTTCGAC TTTATTAGCA GAAGTTAAAC	600
	CATCTCGTTG TCGCGTTGTG TTTGCTAAGA TGATACCGTC AAATGTCTCA GTAATCGCTG	660
15	GTAATAGTGC TTTTAAGCCA TCGAAATCCA TATCAGACGT TAGTTTTAAA TAAATTGGCA	720
	CTGTTACATC ATGTTGTTTT TTAAATGCTG TTAAAGCTTG GCATAACATT GAAAATTCAT	780
	CTTTATCATG GAAGTTTTGA AGATTTTCAG TATTTGGAGA ACTGATGTTG ACTGTGAAAA	840
20	ATGAAACGTC GTGTTTAAAC GTATCAATAA CCTTTATATA ATCTTGATAA CGCGCTTCAT	900
	AAGGTGTCAT TTTATTCACA CCAACATTGA TACCAACAGG TACTTGATAA GCATTTTTAC	960
25	GCAAATGACT TAGTGCTTTG TTCATACCAA TATTATTGAA GCCCATTCTGA TTTATCAAGG	1020
	CGTCATCTTC TAATAATCTA AACATGCGTG GTTGAGGGTT ACCCGGTTGA GGTTTAGGTG	1080
	TGATACCACC TAATTCTAAA GCACCGAATC CAAGGTGTTT CAATGCTTTT GGTACTTCGC	1140
30	AAGATTTGTC GAAACCAGCT GCTAAgCCAA TTGGATTGTC GTACGTATTA CCTTGTATCG	1200
	TTTGTGATAA CGTTGGATTC TTATAAGTAA ATAGTTTATC GACGACTGGG AATAAAACCG	1260
	GaAACTTTTG TaACGTTTTT AATGCATCGA TAGTTAGTCC GTGTGCTTTT TCGGGTTCTGA	1320
35	TTTTGAATAA GAAAGGTTTA ATTAATTTGT ACATGAGTAT GCTCCTATTT CATTATATTT	1380
	GAGGCTTACT ATCCTCAACT TAATATATGT GAAATATATT CTTTTAATAG ACTAGCATTT	1440
	CCATACATAA TTTCTAGTT AAAACTAAAA AGTTTTGAAA ATTGACGCAA gTTTGAATAA	1500
40	CGTTTTTAAG ATTAAATCAT CCTAATTAGG CAATATTATA GTATAAAGTA AGTAGATTGG	1560
	AAGGTGTTTG TATGAATGAA CAATGGTTAG AGCATTTACC TTAAAAGAT ATTAAAGAGA	1620
45	TTTCACCACT GAGTGGTGGT GATGTAAACG AAGCATATCG AGTCGAAACA GATACGGATA	1680
	CATTTTTCTT ACTTGTCCAA CGTGGACGTA AAGAATCATT TTATGCTGCA GAAATTGCAG	1740
	GTTTAAATGA ATTTGAACGT GCAGGTATCA CGGCACCTAG AGTAATTGCA AGTGGCGAGG	1800
50	TTAACGGTGA TGCGTATTTA GTGATGACGT ATTTAGAAGA AGGGGCTTCA GGGAGTCAAC	1860
	GCCAATTAGG GCAACTCGTA GCTCAATTAC ACAGTCAGCA ACAAGAAGAA GGCAAATTTG	1920
55	GCTTCTCATT ACCTTATGAA GGTGGCGATA TTTCTTTTGA TAATCATTGG CAAGACGATT	1980

	GGCTATGGGA TGCCAACGAT ATCAAAGTAT ATGACAAAGT GCGACGTCAA ATTGTGGCGG	2100
	AATTAGAAAA GCATCAAAGT AAACCGTCTT TATTACATGG TGACCTATGG GGTGGTAATT	2160
5	ATATGTTCTT ACAAGATGGT CGTCCGGCGT TATTTGATCC AGCGCCATTA TATGGTGACA	2220
	GAGAATTCGA TATCGGTATT ACAACGGTAT TTGGTGTTTT TACGAGCGAA TTTTATGATG	2280
10	CGTATAATAA ACATTATCCA CTCGCAAAAG GTGCATCCTA TAGACTTGAA TTTTATCGTT	2340
	TATATTTATT GATGGTCCAT TTATTGAAAT TTGGTGAGAT GTACCGTGAT AGTGTGCGC	2400
	ATTCTATGGA TAAGATTTTA CAAGATACAA CAAGTTAGTT AAGACGTTAG ATTGAGATAA	2460
15	ATAGATAATA TGCACAGATA TTTTACAAT GAGAAGCGAT ACAGCTGCCT CAATAAAAAA	2520
	ATTTGTGCGT TTTTATTGTT GGAATAAAA ATTTTAATCG CTATTGTAA TTTCTGTAAT	2580
	GTAAACAAG GTTGAGTTAC AATAAAGTG ATTTTATAAC TTTTGTTCATAATAATTCT	2640
20	AGGAATGATA CATATTTATT GATACAATA TTTTGAATAT AATCATAAAA CAATATTTAA	2700
	GTATAATTGA ATGTTTGAAT ATCATATATT GATACAGTTT CTAATAATTT TAAAATAATT	2760
	TAAATGGAGA GAGGTGTAAA TGATGAGTAC AGTTCAAAGT GATATTTTAA AGACCAATAG	2820
25	TGCATCATCA TCTATTAAAA GCGCTGTTGA AACATGTAAT AATGTGTCGA AACCGGATAA	2880
	AGATGAAAGT ACAACAGTAA GTGGAAATAA TAATGCTCAT AGTGTGATAG ATGATTTGAT	2940
30	GAGTAAGAAT CAATCTGTTG CTGAAGCAAT ACGAACTGCG AGCGATAATA TACAAAAAGT	3000
	TGGTGAGGCT TTTGACCAAA CTGACGTAAT GATTGGTAAT GAAATTGGTA AAAATTAAAA	3060
	CGTGGTGAAA TGATGTCGAA TAACTGGAT GAAATCAATA AAATAATCAC AGCGAAACAT	3120
35	GAGCAAATGG ATGACTTATA TGATGAAAAG CGAGAGGTTA AAGCATTGAT AGATGAAAGT	3180
	GATGCGCTTA ATCATTCGAT AGATCAATTA TATCAACATT TAGGTGAGCG TTATTATAGT	3240
	AGCAATATGG CTAGTCGTAT GGAACAGTTC CGCGATGAAT TTCATTTTGC GAAACGACGT	3300
40	TCAACGGAAG CGTTATACGA GCAGCAACAG CAAATTCAAC ATGGCATTCTG TAAAGTGGA	3360
	GAAGAGATGA TTGACTTGGA AATGCGAAGG AATGTTGAAA TTGAGACGGT GACAAAGGAG	3420
	GAAAATAAAT GGAAACAATA GGAAGCATT TTTATTTAAA AGAAGGTTCTG CAAAAGTTAA	3480
45	TGATTATTAA TAGAGGmCCA aTTGTAGAAA TTGAAAATCA AAAGTATATG TTTGACTATT	3540
	CTGCATGTAA ATATCCGATT GGTGTTGTAG AAGATGAAAT TTATTATTTT AACGAGGAAA	3600
50	ATATAGATTC AGTTATTTTT AAAGGTTATT CTGATCAAGA TGAGGTTAGA TTTCAAGAGT	3660
	TGTTTGAAAA TATGAAACAA AATTTGGATA GTGAAATACA ACGTGGAGAA GTTACACAAC	3720
	AATAAGAAA TACTTTTTCT TTATTGGGGT GGGACGACGA AATAAATTTT GTAAAAATAT	3780

55

	ATGTCATTCA TAATCATTTG AACTAAACGT AGCAGCCTTA AATTTTAAAA AAAGACACAT	3900
	ACCAACTTCC GAAATGTAGA TGAATTCTCT ACAATAACGG AAGTTTTTCT TTTAATATTG	3960
5	AAATTTCTCA AGGATAGGTC TATACTTTAT AAATCGTAAT TATTACGATT TATAATCAAA	4020
	AACAATAACT TGAAATAGAT CATTGAGGGA GTGTTAATAT GCAACATCAT AAAGTGGCTA	4080
10	TTATcGGTGC CGGTGCTGCA GGTATAGGTA TGGCCATTAC CTTAAAAGAT TTCGGTATAA	4140
	CAGATGTCAT TATTTTAGAA AAAGGAACAG TAGGACATTG ATTTAAACAT TGGCCGAAAT	4200
	CGACCCGTAC GATCACGCCA TCATTTACGT CTAATGGATT TGGCATGCCT GATATGAATG	4260
15	CAATTTCCAT GGATACTTCA CCAGCATTTA CATTTAATGA AGAACATATT TCCGGAGAAA	4320
	CATATGCTGA ATATTTACAA GTGGTTGCCA ACCATTACGA GCTGAATATC TTTGAAAATA	4380
	CAGTTGTCAC AAATATATCT GTAGATGATG CATATTATAC GATTGCAACG ACAACAGAGA	4440
20	TATATCACGC GGATTATATC TTTGTGCGAA CAGGTGATTA TAATTTCCCT AAAAAgCCAT	4500
	TTAAATATGG TATTCATTAT AGTGAAATTG AAGACTTTGA TAACTTTAAT AAGGGGCaAT	4560
	ATGTGGTTAT CGGAGGTAAT GAAAGTGGCT TTGATGCTGC ATATCAACTT GCAAAAAATG	4620
25	GCTCTGACAT CGCACTTTAT ACTAGCACAA CCGGTTTAAA TGATCCGGAT GCTGATCCTA	4680
	GTGTTAGATT GTCACCTTAT ACACGTCAGC GACTAGGTAA TGTCATTAAG CAAGGTGCTC	4740
30	GCATCGAAAT GAATGTACAT TATACAGTTA AAGATATTGA TTTTAACAAT GGACAGTATC	4800
	ATATCAGTTT TGATAGCGGA CAAAGTGTGC TTACACCTCA TGAACCAATA CTAGCAACTG	4860
	GCTTTGATGC AACAAAAAAT CCAATCGTTC AACAAATTATT TGTGACAACA AATCAAGATA	4920
35	TTAAATTAAC AACACATGAT GAATCGACAC GTTATCCGAA TATTTTATG ATTGGTGCAA	4980
	CAGTTGAAAA TGATAATGCC AAATTATGCT ATATCTATAA ATTTAGAGCG CGATTTGCAG	5040
	TACTTGACACA TCTTTTAACA CAGCGGGAAG GcTTACCAGC TAAACAAGAT GTCATTGAAA	5100
40	ATTATCAAAA AAATCAAATG TATTTAGATG ATTATTCATG TTGTGAAGTG TCATGCACAT	5160
	GTTAGAAGTG AAATATGATA TGAGAACTGG GCATTATACG CCCATACCTA ATGAACCTCA	5220
45	TTATTTGGTT ATTAGTCATG CGGATAAACT TACCGCAACA GAAAAAGCGA AATTAAGATT	5280
	ATTAATCATA AAACAGAAAT TAGATATTTT ATTGGCAGAA AGTGTAGTTT CTTcGCCTAT	5340
	AGCGAGTGAA CATGTGATAG AACAAATTGAC ACTATTTCAA CATGAGCGAC GACATTTAAG	5400
50	ACCTAAAATA AGTGCGACAT TTTTAGCCTG GTTGTGATA TTTTAAATGT TTGCATTGCC	5460
	AATCGGTATC GCTTATCAAT TTTCAGATTG GTTTCAAAAT CAGTATGTGT CAGCATGGAT	5520
55	AGAATATTTA ACTCAAACAA CATTGCTCAA TCACGATATA TTACAGCATA TATTATTTGG	5580

	ATTGATTAGT TTATCAACTG CTATAATTGA TCAAACAGGA CTCAAATCAT GGATGATATG	5700
	GGCAATTGAA CCGTCAATGT TATGGATAGG ATTACAAGGT AATGATATCG TGCCACTATT	5760
5	AGAAGGGTTT GGATGTAATG CAGCAGCTAT TTCACAAGCA GCACACCAAT GCCATACCTG	5820
	CACGAAGACA CAGTGTATGA GTTTAATAAG CTTTGGTAGT TCTTGTAGTT ATCAAATAGG	5880
10	TGCGACATTA TCTATTTTTA GTGTAGCTGG AAAGTCATGG CTATTTATGC CGTACTTAAT	5940
	ATTAGTACTT TTAGGTGGCA TCTTACATAA AGGATATGGT TGAAAAAGAA TGATCAACAA	6000
	CTTAGCGTTC CGCTACCTTA TGATAGGCAA TTACATATGC CAAATATACG TCAAATGTTG	6060
15	CTACAAATGT GGCAAAATAT ACAAATGTTT ATCGTTC AAG CGCTACCTAT TTTTATCACA	6120
	ATCTGTCTTA TTGTTAGTAT TTTATCACTA ACGCCAATTT TGAATGTTTT ATCACAAATA	6180
	TTTACACCTA TATTATCGTT ATTAGGCATC TCGTCAGAAT TGTCAACCAGG GATTTTATTT	6240
20	TCAATGATTC GAAAAGACGG CATGCTCTTG TTTAATTTGC ATCAGGGCGC CTTATTACAA	6300
	GGAATGACAG CAACACAGTT ACTACTACTT GTGTTTTTTA GTTCAACATT TACAGCGTGC	6360
	TCGGTCACAA TGACGATGCT TTTGAAACAT TTAGGTGGTC AGTCAGCACT AAAATTAATT	6420
25	GGAAAGCAAA TGGTGACATC ATTGTCTTTA GTTATTGGTG TAGGCATCAT TGTAAAATA	6480
	GTAATGCTGA TTATTTAAAA AAAATGAACT ATAAGTGAAT ATAGAGTCAT GTCAGTCAAT	6540
30	AGGAGATCTA TCTTGGAATA TGCTATTTCAT ATGAAGTATA AGAGGAGAGT CGCAGATGAA	6600
	AATAGTTATT ATAGGTGGGT TTTTAGGTGG CGGTAAAACG ACTGTCTTAA ATCATTTGCT	6660
	CGCTGAATCA TTAAAGGAAT CGCTGAAACC AGCAGTCATC ATGAATGAAT TTGGGAAAAT	6720
35	GAGTGTGAT GGTGCCTTAG TATCTGAAGA CATACCTTTA AGTGAAGTGA CAGAGGGGTG	6780
	TATCTGTTGT GCAATGAAAG CAGATGTATC AGAACAGTTA CATCAATTAT ATTTAAAAGA	6840
	GCAAACAGAC ATTGTATTTA TTGAATGTAG TGGGATTGCA GAACCGGTCT CTGTCTTAGA	6900
40	TGCTTGTTTA ACGCCTATTT TAGCTCCGTT TACAACAATT ACACATATGA TTGGTGTAAT	6960
	AGACGCAAGC ATGTATAAAC ACATTAAATC ATTCCCTAAA GACATCCAAG GCTTATTTTA	7020
45	TGAGCAATTA GCATATTGTT CTGTCTTATT TGTTAATAAAA ATAGATTGAG CAGATGTTGA	7080
	AACAACGAGC AACTATTGA AAGATTTAGA AGTTATTAAC CCAGAGGCCG ATATACAAGT	7140
	CGGTATGCAT GGCAGCGTCA CTTTGCCAAT ATCAGTTAGA CAAATGACAG CAACTTCTGA	7200
50	CAATAAACAT AAGTCTTTAC ATCAAATGAT TAATCATCAA TTTGTGCAAT CACCAGTCAA	7260
	ATGTACTAAA GCAGAGTTTA TAAAACGTTT AGCATGCCTT CCGTCTCATA TTTATAGGTT	7320
55	GAAAGGGTTT ATGACATTTG AAGACACCGC ACATACGTAT CTCATTCAAT TTACACAAGG	7380

EP 0 786 519 A2

	CGGAAAGGGT	ATTTCAAAAG	AAGACTATCA	ATGTTTGGAA	CAGTAGTGTT	TTCAGTGGAA	7500
	GAGAATGGTT	AACATGCCTT	CATGTATAAT	AACGAGTTGA	TTTGAACGTT	TAAGCGTAAA	7560
5	TAAAAATAAG	CTTGGTCAAG	CATCAAATAT	AATTTGAAAA	CTGTCCAAGC	TGTTTTATTA	7620
	GAGAACAATC	AATTAACCCC	ACATATTTAA	TAATACATCA	GCAAAGCCTT	CAGGTTTTTG	7680
	AATATAACCT	AAGTGACCGC	CTGGAATATC	TACAATAGGT	ATGCCAGTTT	CTTTATTTAT	7740
10	ATAAAAGTTA	ACATCTTGTT	GGAAGGAGCC	TCTAGAATCT	GTCCCATTTA	GTAGGGTGAT	7800
	TTTATCGCTG	TATTTTGTGA	AATCATCCAA	AGTAATATCT	GAATGCGTAT	ATTGTCTAAT	7860
15	TTCAAATTCT	GACCAGAACA	TCGTACGTTT	GTACTGTTCT	ATACGTCCTT	CTTCAGTATC	7920
	AGCAGGTTGA	GACATCATTT	TTGCATCAAT	TGGTGCAGTA	TTTAATGTTT	CGCCAAATGT	7980
	TTTCATGCCT	TTTTCTAAGC	CTTCTGTTAA	AATTTGATGC	ACAATGTCAT	CATTTTTATC	8040
20	TTTCCAATAA	GTACTGTCTG	GTAAAAATGT	ATTAATTGGT	GGTTCGTGAA	ATGCAATCTT	8100
	TTTAACGACT	TCAGGGTAAT	CTTTTAACAC	ATGCATCGCA	ACGATTGAAC	CTGAACTTGA	8160
	ACCTAATATA	TAGACAGGTT	CATCACTTAA	TGACTTTGCA	AGTTCGGCAA	TGTCCTGTGC	8220
25	GTGCGGTTTG	ACACGATAAT	CACTGTCAGG	GTTTGAAGCG	GAATCAGGGA	GTGGTTCAGT	8280
	TAACGCTGCT	TCTCCATAAT	CACGACGATC	AACGGCTACA	ACAGTAAAAT	GGTCTTTTAA	8340
	CTGTTCTGCA	AGAGGCAGAA	AAATGTCTCC	GGTACCGTTT	GCACCAGGAA	TAAAGATGAG	8400
30	CACGGGTCCT	TGTCCGACTT	GGTGGTATCG	TAATTTAGCG	CCTTGTAATT	CTAAAGTTTC	8460
	CATATTCAAT	GACCTCCATT	TGTTAATTGT	TAGGTGATAA	ACCTAATAAT	TAGCACCATT	8520
35	TTGTATAACT	TATTTTCTCT	TTTCTTTCAT	CTGTTAAACC	CAGTTCATCT	AAAAATACAC	8580
	CTAATTTTTC	AGGCTCAATA	TATGGATAAT	CAGCAGCATA	AAGAATTCTA	TCAATACCTA	8640
	CTTCTTTCTT	GACTAAATCA	AACTGTGGCT	TCGTTAACAT	GCCACTCGGT	GTGATATAAA	8700
40	AATTATTTTT	AAAGTAATAG	CTTACAGGGT	GGTTCAAATG	TTCAGCGAAT	AAAGCTTCAT	8760
	CCATACGTTT	TAAGAAGAAT	GGGATAAACT	CACCCCAATG	TCCAATAATC	ATATTTAACT	8820
	TTGGATAACG	ATCAAAAATA	CCAGATAATA	CTAGATGTAT	TGTATGAATG	CCGACATCAA	8880
45	TGTGCCAACC	ATAACCAAAA	CAAGCAAATG	TTGCCGCAGT	TACTTCAGGA	TAATTTCTCT	8940
	TATAGTATGA	TTGATAAATG	TCACTGTTAA	CTGGCGCGGG	ATGTAGATAA	ATCGGTACGT	9000
50	CTAAATTTTC	AGCTGTTTTG	AAAATAATGT	CATATTTGTC	TTGATCAAGA	AAACCATCTT	9060
	GTGCACGTCC	CATAATGAGC	GCACCTTTGA	ATCCTAAATC	ATTGATGCAA	CGTTCGAATT	9120
	CTCGCGCTGC	GGCTTCAGGC	TCATTGATAG	GTAAAGTTGC	AAAGCCTACA	AAGCGATTGG	9180
55							

	TCTGACCAAC	CAAATTTGAA	GGAGAACCAT	TTCCATAAGA	TAAGACTTGA	ATTTGAACGT	9300
	CTTGATTATT	CATAAATTGG	ATACGTTTCA	TATGATGTGA	TAATTCGTCG	GCATTTGTAA	9360
5	AACCTGTCTT	TTTTTcAAGG	CCTTCTAACA	TTACTTTTCAT	CGGTACACCT	TTAGGATCTG	9420
	CTGATATCGC	ATTCATCGTT	TCTTTTTGAA	TATCTTCAAT	GACATAATGT	TCTTCAAACG	9480
10	TAATACTTTT	CATTTACTTC	GCCTCCATAT	TGTATTGCAT	GTTTATTGCA	TCTATTGCAG	9540
	AAGCATTTTT	TATATACCTC	TAATTTCAAT	GTTTGTAACA	TAAAATTGAT	CTACCAAGGC	9600
	ATCTCTCCAT	CGCCATTAAT	AAATGTACCT	GTTGGGCCAT	CTGCACCAAT	CGTTGCTAAT	9660
15	TGAATGATTG	GCTTGATTCC	TTCAGAAAACG	TGTTTGGAAT	TATTACTAAA	ATCACCAACT	9720
	AAATCAGTAT	TTGTAGCGCC	TGGATCAGCA	GCATTGATTT	GCATGTTAGG	TAATCCTTTA	9780
	GCGTATTGTA	GCGTTAGCAT	TGTTACTGCC	GATTTAGACG	AACAATAAGC	TAATGAATTC	9840
20	ACTTTAGATT	CAGCTGTTTC	GGGGTTTGTA	ACCATTCCAA	ATGAACCTAA	ACCACTTGAT	9900
	ACGTTGACGA	CAACAGGTTG	TTCAGATTTT	TCTAAGAGAG	GGACGAATGT	ATTCATCATT	9960
25	CGTACGATAC	CGAATACATT	CGTTTGATAT	ACTTCTTCAA	CGTCACGAGG	TGTCAATTTG	10020
	GAAGTGCTG	AAAATTGACC	AGATATACCT	GCATTGTTAA	TGAGGATATC	AAGACGGCCT	10080
	TCTTTTTTCAG	CAATCATGTT	ATAAGCATTT	TTGACTGAGT	AGTCACTTGT	AACATCTAAT	10140
30	TGTACATAAT	GAACACCTAA	TTTTTGTTGAT	GCTTGTTGTC	CTCTTACATC	ATTCCGAGAA	10200
	CCTATATAAA	CTTTGTAACC	CAATGCTTTA	AGTGCCTCTG	CACTTGCATA	GCCTAACCCCT	10260
	TTATTGCCTC	CTGTGATTAA	CACAATTTTA	GTCATTACGT	CCCACCTCAT	CTAAATAAAT	10320
35	GTTTAATAAA	TAATTTCTGT	ACGCTTCAAT	TGAAATATGG	CGATGCTCTA	TTTGAAGGC	10380
	AAATACACTA	GTTGATAATG	ATTGCAACAG	CATATCTGTT	TTGAATTCGT	GTAAGTGTCG	10440
	TCATcGCTTT	TAAATAAGTC	ATAATAAAAA	TCAAATAATT	CTTGATAAAA	TGCGCTTTGG	10500
40	TAAAAACGTA	ATTTATTGTT	GCCTGCTTCA	ATACATTGCA	GTAGTGCCTT	ATTATCGATT	10560
	TTAAATTGTA	AAAGATAATC	TAACGACACT	TGCATAACCT	CATAATTAGA	ATGATAGTCA	10620
45	TCTTTAATTT	GCTTAAAAATG	AGTGATAAAA	ATATCAAGGT	CTCTTTGTAT	GACGTAGTAG	10680
	CATAAATCGC	TTTTATCTTT	GAAATGTCGA	TACAATGTCC	CCATACCGAT	ACCTAGTTCT	10740
	TTAGCAATAC	GATTCATACT	AATGTTTTCA	ACGCCTTCTT	CATCAAAAAG	TTTGTGCGCT	10800
50	ATTTCTTCAA	TTCGTTGCCT	ATTCTCTTTT	GCATCTTTTC	GCATGATTAC	ACCTACTTAA	10860
	AATTCTCTAA	AATTGACAAA	CGGATAACTC	TCCGTTTATT	ATAAAACGTG	TTAAGAAAGT	10920
55	TAGCAATGAA	TTTGCAATAA	CTATTAAATA	TCATAAAAGA	AAAGAGTGTT	GATAATGTCT	10980



ACCTTATCGG TTCAAATGAT TGCTGAAAAA CTGAATGTCA CTACAGAAGA TGTGGAAAAA 11100  
 GTATTAGCTA TGACAGCGCC ACTAGGCATT TTTAGTCATC AATTACAACG ATTTATTCAT 11160  
 5 TTAGTATGGG ATGTCAGAGA TGTAATAAAC GACAATATTA AAGGAAATGG ACAAACACCA 11220  
 GAACCATATA CGTATTTAAA AGGTGAAAAA GAGGACTATT GGTTTTAAAG A 11271

## (2) INFORMATION FOR SEQ ID NO: 12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6261 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

20 CAACCCGTTT AGAACAAAAT AAAAACCGTA CAATTTTATC ATCTTAATGA TTATTGTACG 60  
 GAAAAACTTT TTTACATCAT ATCTGCATGT GCATAATCGA TATCGGTAAA TTTATTATAT 120  
 TGTTCATAA AATGTAACCT AACTGTGCCT GTTGGACCGT TACGTTGCTT AGCAATGATA 180  
 25 ATTTCAATTT CACCGTTTTT ATCATTCGTT TGTGGCTCGA AACCACCATC ATCGTCATCA 240  
 TCTTCATCGC CGCCACGGTT ATAGTAATCA TCACGGTATA AGAATGCAAC GATATCGGCA 300  
 TCTTGCTCAA TCGAACCAGA TTCACGAATA TCACTCATCA TTGGACGTTT ATCTTGTCGT 360  
 30 TGTTC AACAC CACGAGATAA CTGACTTAAT GCGATAACTG GACATTTTAA TTCACGGGCT 420  
 AATGCTTTTA ATGTACGAGA GATTTTCAAG ACTTCCTGTT GTCTGTTATC GGACGCACGT 480  
 35 GAACCACTAC CTTGAATCAA CTGTAAGTAG TCAATCACAA TCATGTCTAA GCCATGTTCT 540  
 TGCTTTAATC GACGACATTT AGAACGTAAA TCATTAATTC GAATACCCGG TGTATCATCA 600  
 ATAAAAATCT TCGTACGTGA TAATTTACCT ACCGCTATAG TAAAACGACT CCAATCTTCC 660  
 40 TCAGTCATAG TACCCGTTCT TAAGCGGTTT GAGTCAACAT TTCCAGAACT ACAAATCATA 720  
 CGTGTGGCTA ACTGATCAGC ACCCATCTCT AGCGAGAAAA TACCAACTGT ATACATATCT 780  
 TCATGCGTTG CAACTTTTTG TGCAATATTA AGTGCGAACG CAGTCTTACC TACAGATGGA 840  
 45 CGCGCTGCAA GGATAATTAA ATCATTTTCG TTGAACCCTG CTGTCATTTG GTCTAAATCT 900  
 CGATATCCTG TAGGTATACC TGGTGTGTTGA CCACTATTTT GATCAAGCTC TTCAGCTGTT 960  
 50 TCATACACTT GTCCTAAGAC GTCTCGAATG TCTTTAAAGC CATCGCTTTC ACGAGAAGAT 1020  
 GATAGCTCTA AAATTCGACG TTCTGCATCA CTTAAAATCG CATCTAGTTC AAGTTCATCA 1080  
 TTATATCCAT CATGGCAAT ACTATCTGCA GTTTGAATCA ATCTACGTTT TAATGCATGC 1140

EP 0 786 519 A2

	TCTGCAAGAT ATTGCGGGCC ACCCGCTTcA TTCAACGTAC CTTCCGTCGA TAATTGATCC	1260
	ATCAATGTTA CAACATCAAT TTCTTTATTA TCITTCATTTA AGTGCATCAT TGCACGGAAA	1320
5	ATATGTTGAT GGGCACCCCT ATAAAACGAC TCAGGAAGCA AAACCTCCTG AGTAGTATTA	1380
	ATCAATTCTG GATCTATAAT AATTGAACCT AAGACAGACT GTTCAGCTTC ATTGTTATGC	1440
10	GGCATTGAT TTTGCTCATA CATTCTATCC ATGAATGGTT ACACCTCTTA TTTCAATCCA	1500
	ACTTTATTGT TCAACTGTGT GTACGCGAAT TGTACCTTCA ACTTCTTTAT CTAATTTAAC	1560
	AGGTACATTC GTATATCCTA GGAATGAAT TCCATTGGT AAATCCATTT TACGTTTATC	1620
15	AATTTTAATA TCATGTTGTG CTTTTAGTGC TTCGGCAATT TGTTTTGTAC TTAAGTACCC	1680
	AAACAATTTA CCACCTTCAC CAGTTTTTGC TGaTACTTCA ACTTCAATGT TTGATAACGT	1740
	TTCTTTTAAT GCTTTAgCAT CTTCAATTTT TGTTGGCGT TCTTGTTTTG CACGTTTTTT	1800
20	CTGTAACTCT AATTGTTTAA GGTTACCTGG TGTTGCTTCT ACAGCATAAT TCTTTTTCAA	1860
	TAAGAAGTTA TTTGCATAAC CTAAGGTGAC TTCTTTAACT TCACCTTTTT TACCTTTACC	1920
	TTTACCTTTA ACATCTTGTTG TAAAAATTAC TTTTCATGCAT CTTCACTCCT ACTTAATTGT	1980
25	TCTGTAATTG CTTGTTGTAA TTGTGCTATC GCCTCTTCGA CTGTCACACC TTTAAGTTGT	2040
	GTTGCCGCAT TGGTTAAATG TCCACCGCCA CCAAGTGCTT CCATTGTTAA CTGGACATTT	2100
30	ACTGAACCGA GTGAACGCGC AGATATACCA ATCAGATTAT CTTACAGTCT CGCAACAACA	2160
	TATGATGCTT CAATACCTTC TAAACTTAAC AGTTCATCTG CTGCTTGTCG AACTGTTACT	2220
	GGATGATAAA TTTTATCGTC TGAACCATGC GcAATGGCTA TGCCATTATC TTCAACTTTT	2280
35	ACAGTTCGAA TTAATTCAGA TCGATTAATG TAAGTATCCA CATCATCTTT TAAGAAATGT	2340
	TGCGTTAAAA TCGTATCTGC ACCATGTGCA CGTAAATAAC TCGCTGCATC GAATGTTCTT	2400
	GATCCTGTTC GTAATGTAAA GTTTCTTGTA TCTACAATAA TACCTGCATA CATCACTGTT	2460
40	GATTCAAGAC GTGTTAAACG TTGTTCTGTT GGTGATATT CCAGTAACTC TGTACCAAT	2520
	TCAGCTGTCG AACTTGCGTA TGGTTCCATA TATATCAACA ATGGATTAGA GATGAAGCTT	2580
45	TCACCACGTC TATGATGATC GATAACAAC TTAGCGTTTG CTTTATTTAA GACATTTTCA	2640
	TCTAAAACCA GTTCCGGTTT ATGCGTATCA ACAATCACTA CGGTTGTCTT AGATGTCATC	2700
	ATATCCCAAG CATCATCTGA TGTAATAAAT CGCTCTCTTA ACTCTGGCTT TTTATCTATT	2760
50	TCGTTTCATCA CGCGTCGTAA TGTTGGATCA ATGTCAGTCT CATTTAATAC GATGTATGCT	2820
	TCTAAATTAT TCATCATTGC AAATCTAGAC ACACCGATTG CTGCACCAAT TGCATCTAAG	2880
55	TCAGGACGTT TATGTCCCAT GATAATGACT TTGTCACCCT CTGCAAGGAT ATCTTTTAAAC	2940

	CCATAGAAAC GCACATTACC ATTAATACTT TTAATTGCAA CTTGGTCGCC ACCGCGTCCT	3060
	AATGCTAAGT CTAGGCCTGA TTGTGATAAT TCACCTAAGT CGATTAAATT TTCAGTACCT	3120
5	TCACCAACAC CGATACTTAA TGTTAATTGG GCACGATAAC CAACACTTTT TTCACGTAAT	3180
	TGACTCAAGA TATCAAATTT AGATTCTTCT AAGTCAGCTA ATATTTTTTG ATTTAAATAG	3240
10	GCTACGAATT GATCGGAACT GTATCTTTTG AAAAATATAT TATACTCAGT TGCCCATCGA	3300
	CTAATGACAC GCGTTACCAT TGAGTTGATT TCCGAACGCT GCGTATCATT CATATTTTGC	3360
	GTAATCTCAT CGTAGTTATC TAAAAATAAT GTCGCAATGA TTGGTTTAGA ATTTTCATAT	3420
15	AGTTCATTTG TTTGTACTTG TTCAGTTATA TCAAAGAAAT AGAGGCAGTG ATCATTCTCA	3480
	GAATAACGTA CTTGGAAATG ATACTGATTA TATTCTATTT CAACGGATTT CACTCTATCT	3540
	AATTGCTTTA AAATGTTTGG AAATACTTCA TTTACAGATT CAGAAATGAC ATTCGCTTCC	3600
20	ATATGATCTG TCATAAATTG GTTAACCCAT TCGATGTGAT CATTTTCATC TAAAACAATG	3660
	ATACCAATTG GTAAATGTTT GATTGCTTTA TTATTTGTTG TTGAAATTTG AGCACTCAAA	3720
	CCATCTACAT AACTATCCAT TTTCATTAAA GCTTGTCTGA ATAAATGAT GCTAACAATA	3780
25	ATCATCACGA CAAGAACGAT AGATGCAATT AGTGCTATAA GACTATTAAA GATAAACCAT	3840
	ACACCCATTA AAACAATTGC TGTGATGATC ATGATGACAA ATGGTATTAG TAAAGCTTTC	3900
30	TTAGTGGA CTGCCGATT CATTTCCACCT CTATTCACCT TTTAGAATTA TTTTTCATGA	3960
	TTGCTTCAA ATTCAAACCT AAATCGATAA CACCAAGTAG TCCTACAATA TGTGTCGTAG	4020
	GTGTCAGTAT TGTACCGATA ACCAATAGTA AAATCGTTAC TGCATTCGGC AAACCTTTCG	4080
35	CTTTACCAA GAAATGAATA AACTTTAAAC CTTGAATATA CATTACTAAT GATAACACAA	4140
	GTTGGAAGTT TAAAAGAATG CTCTGGAACA CACTCGGTTG ACCTGTAAAT AATAACATA	4200
	TGATAACAAT AATGTATATC CATAATAAAA TACCGCTCAT TTGCCACGCG AAAAGTGGCT	4260
40	TAAATACAGG TGTAGCGATT TTAAATTTTC GTAAAATCGG AAATGTAACG ATTAAGTTAA	4320
	TTAAGACGAT TAAAAATGTA ATGATAATGA TGAAACCTGG TAATTGAACG GTCGCTTGTC	4380
	TAAACCTTTC TTCTAATATT TGGGTCATAT TCGCATCGGC ACCGCTCATC GTAATCGCTT	4440
45	CATGTAATGT TTGCTTGAAA GGTTTTACTA TGCTCGCTGA TGGTGGAATC CTTCCGAATG	4500
	TTTGTAGTAA CATAAAAGCG ATTAATGAAA TTAACTCAT CGCTACTGTT GTTACGTATA	4560
50	ATATTCTTTC TTTAGACGTT CTTTCTTTGA GCAATTGACC AATAATTAAA CTTGCAATTA	4620
	AGACTAATAT GATGGCACTT AAAACGAAAG TATTACCTAA AACAGTTGTT ATAATTACTG	4680
	TAATAAGTGC ACTAATCCCG AAAGATTGTA TTGATTTATT CCATAAAACG ATACCTGGTA	4740
55		

	CAAATACCAA CGCAATCGTT GCAATTATTG TTGCTTTAGG TTGTATTTT GAAAACACAT	4860
	AAGCCACTCC CATATTTTAA ACTATAGCTA TTATTTTAACT CTCTTTAATG AAAATTAACA	4920
5	ATTATAGAT TGTATGCTTC TATTTCAATTT AATTGAATAA TAACTTTCAT GTTTTATAAG	4980
	TAATTAACAT ACTCATTGGA ATCGCTTTTG TGTGCTTTCA TTTTCAACAT GATTATTTAA	5040
	TCCCCTACA TAGCAATCAA GCTTGATTTA GATTTACAAT ACATTTCCAC TCTCATGTAC	5100
10	TCTAGATGTT TTTGAATATG ATAACTGTGA TTTAGTGGCT TCATTCTTTG AAAATATATA	5160
	TTATTACTTA CGCTTAAAT GCTTTAAATT TAAGAAATGA TATAAGTTAG GTGCCCAGGT	5220
15	ACTAAAGTTT AGTAGGATC CATCATGCCC AACATTATCA GGCACGAAGA AATGACGATG	5280
	ATATTTAAAA CGTTCACCTA ATGCACGAAC TTGATCATCC GGATATAGCA AATCATCTAT	5340
	GAACCCCATC GTTAACACTT TTGTTTCTAA ATTTTAAAA ACATGCGTTA CGTCTGTGCG	5400
20	ACCTCGGTCA ATGTTGTGAC TATCCAATAC ATCTAGCAGT GTCAGATAAC AATTCAAATC	5460
	AAAATGTTCT TTAAATTTAT TACCTTGATG TTGTTGGTAT GCGACTACTT CATCCGGCGT	5520
	AAAACGTTCA TCATAACTTT TTGATGATCG ATATGTCAA AAACCTAATT GCGGTGCAAT	5580
25	ACTTAGACCT TCCTTACCAC CAAGATGAAT GGCTTGCCCT GCAATTTTCAT TGAAAGCTCT	5640
	ACTATAAGAT GATGTTGAC TTGTTGCAGC AAGGATAATG GCTTTATCTA CTTCAAACCTG	5700
	TTGATTGTAG AGTAGTTCCA TTGCTTGCAT ACCTCCAAGA CTTCCCCCTA TTAAATATT	5760
30	AATCTTATCA TAACCAAGGG CTTGTATACC TCGTTCATTC GCTCTGACTA TATCTCTTAA	5820
	TGTTAATTTT TTAGGAAAT GAGGGTCGTT TAAAGGTGAA CTTGAACCGA AAGGACTACC	5880
35	AATAACATCA AATGTTAAAA ATTGATAATC GTGAATGGGT ATATATCCCC CATCAATAAT	5940
	TTCTCGCCAC CAACCCGGAT AATCATCTGT TCCATATGTT AAATGATTGC CAGTTAATGC	6000
	ATGACAAACT ACAACTAATG GTTGTCCATG ATAACCGACA TGCTCATATC TCAAACGCAA	6060
40	GTAATCTATG ACTTCCCCAG ATTCTGTAAT AAATCCCCT AAATTTAAAG TATCTACTGT	6120
	GTAATTTGTC ATTGTTCTTT CCTCCTTAAA CAAAAAACT TCTCACCCTA TTGAAAAGTA	6180
	AGAAGTCTTT ATACTTATCA TTCGAGTAAC TCGTTGGTTT TAGCACCGTG CTATAAAGTC	6240
45	GGTTGCTGAA GTATCACAGG G	6261

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1222 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

5 ATGCGATTAA CTCTGGAAAT ATCTTTTCCA TATTTACGTh TTAAATTATT CAGCAAATTC 60  
 ATACGAGaTT CATACTCGTT yAACTTGT TCGTCGAATT CTGTATTAGC CATTTCATCA 120  
 TATAACTCAT GTTTTGCATC TTCTAAAATG TAGTAAAATT GATCAATATC TTCTTTTAAT 180  
 TTGTCATATT TGTTTGGAAC TATATCGTTT ATTGTTAACA AATGGTTGCT TAGTTCATAT 240  
 10 AAACGATCAG TGATAGCATT TTCATCCGTT AATGTCATAT ATGCGTTATT AAGCGCTAAG 300  
 CTTAATTTTT CAGAGTTTTG AATGCGTTA ATATCTATTT CAAGTTGCTC TATTTGCGCT 360  
 TCTTTTAGAT GTGCTTCAGA CAATTCTTCT AATTGGAATT TCATTAAATC TAAACGCTGT 420  
 15 AGCAATGCTT GGTCTGCTGA TTCTAAATCT TCTAACTCTT GCTTTTGGC TTTATAATTT 480  
 TGAAAAGTTT GGTGATATTT ATCCAACAAA TCTTGATAAC GTGATTCTGC GTAATTATCC 540  
 20 AATAATGTTA AATGGTATTT TTGTTTCAAC AAAGACTGCG TTTCATGTTG GCCATGAATA 600  
 TCTAATAATT CTTGCATAAC TTTTCGTAAA TCTTGTAAG TAACTGTTTG ATTATTAATT 660  
 TTACAAAGAC TTTTACCAGA GCTGAAAATT TCCCGTTTAA CTAATAAAAA ATCTTCATCT 720  
 25 ACATCAATAT CCATATTTTT CAATATATGT ATAGCATCTT TACTCTCGTC AATATCAAAT 780  
 ATACCTTCGA TGACAGCCTT TTTTTCACCA TGTCTTACAA AATCAGATGA AGCTCTCATT 840  
 CCAATTAATT GTCCAATTGC ATCTATAATA ATTGACTTAC CTGAACCCGT TTCACCACTT 900  
 30 AAAACAGTTA AACCATCAGA AAATTGAATT TCTAATTCTT CAATAATAGC AAATTGCTTG 960  
 ATTGATAAGG TTTGTAACAT AAATCATCG CATCCTTATA ACAAATTGAA AATTCTTGAC 1020  
 35 TTGATTTTCA CACTTGCCCTC TTTGCTTCGA CAAATAATTA AACAAGTATC ATCACCACAA 1080  
 ATTGTGCCTA GACTTCTTC CCAATTGATT TGGTCTAATA TAGCTCCAAT AGATTGTGCA 1140  
 TTACeAGGTA TGTTTTTAGA ACAAGTAAAT TATCAGTACC ATCTATATTA ACAAAGGAAT 1200  
 40 CCATTAAATA ACGTCCCAAT TT 1222

## (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1021 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

55 TTGTTATTA TTACnTnAAA TAATTGCATT ACTTTTTACT GATGGTACAA CTTTCCATCC 60

TTCTTTTGGC ACGACATAAT TGTCTTTATC TTGAACTAAA TATCCGCCAG ATACTGAAAC 180  
 AAACTCTTCT TCGTTACTGT CTATAGTCAT ATCAATTCTT AATAATCTTA CATTCTTCTT 240  
 5 TTGTTTTAAA ATATCTAATG CTTTCATCTGT AAATTTTGGT GCAATAATGA CTTCCAAAAA 300  
 GATACTATGC AATTGCTCTG CTAACTCAGG TGTTACAGCT CGGTTTAATG CAACAATTCC 360  
 ACCAAATATT GATTGACTAT CCGCTTCATA CGCATGTTGA AATGCTTGTT CTATCGTGTC 420  
 10 ACCGATACCA ACACCACATG GATTTCATGTG TTTAACCGCA ACTGTAGCAG GTGTATCAAA 480  
 CTTTTTAACT AAAGCTAGTG TAGCATCTGC ATCTTTAATA TTGTTATAGC TTAATTGTTT 540  
 15 CCCATGTAAT TGTTTAGCGC CTGCAATCGT GTGCTTAGCA TTCGAAGTTC TCACAAAATA 600  
 CGCTGATTGT TGTGGATTTT CTCCATATCT TAAAGTTTCT TTATCCCCTT TAAAGAAACG 660  
 TACAATCGCT TCATCATATT CTGCAGTATG CTCAAAACT TTAATCATT AATGATTGCT 720  
 20 ATATGACTCA TCTAACGAAT CGTTTCTTAA TCGCGTCAAT ACTTCTTGAT AATCTGCCGG 780  
 ATGTACAATT GTTGTTACAT GTTTATAGTT TTTAGCTGCA GCACGTAACA TTGTTGGACC 840  
 ACCAATATCA ATATTTTCAA TTGCTTCGTC CATCGTCACA TCAGGGTTTG CAACAGTTTG 900  
 25 TTGGAATGGA TATAAATTAA CTACTACCAT ATCAATTAAA TCTATATGTT GTTCTGATAA 960  
 TTCATTTAAA TGCTGCGGTT TATTTGATC AGCTAAAATG CCACCATGAA CAGCCGGATG 1020  
 T 1021

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3759 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TCATTCACTC CTAAATTGTT ATTACACTAT TACACaTAGC TAATCATCAA TGTGAAATCA 60  
 CCTTCAAAGA CACTATCCAA ATCTTCAGAA GTCAAAATAA AGTTTGTAAC AGTAGTCAGT 120  
 45 TTGAAAATTT CACCATCGAC AATCATTGTC CCTTCGCCCTT CCAACACTGT AACTAAACAG 180  
 AACTCTCTAG GCTTCATATA ATTTAACGTG CCAGAAATTT CCCATTTAAC CAATGTAAAG 240  
 AAATCATTCTG ATACAATGTG TGTACACTTA TGGTTTTCAA TAATTTGCTT TTCAGGCAAA 300  
 50 ATATTAGGTA ATGGTGCATT GACTGAATA ACGTCTAAAG CTTTTTCAAT ATTTAACGGT 360  
 CTATCATTAT ATTGATTATC TTGACGATTG AAATCATAAA GTCTATATGT AATGTCTGAC 420

	ATAAAAtAGa ATTCyCCAGG kTTTACtTTA AtatATCyAA gTAtCGaCtC tATCGTTCcG	540
	TGTTGAACAT GATTGCAAC TTCTTCTCTA GACTCTGCTA ATGTCCcAtA AACTATTTCT	600
5	GCATCTTCTT CTGCATCTAT AATATACCAa CATTcAGATT TGCCATATTG CCCgTTTTCA	660
	TGCTCATAAG CATAAGAATT ATCAGGGTGC ACATGAATAG AAAGTGATTc TCTTGcATCC	720
	ACTATTTTtAG TTAGAAGCGG AAAATCTTTG CTTGGGAAAT CACCAAAcAA TTCACGATGT	780
10	TCTGACCAAA TACGGTCTAA TGTTTGACCT TGATATGGTc CATTaATAAT CTCGCTCGTA	840
	CCATTTGGAT GTGCTGACAC ACACCAACAT TCCCCAGTT GTATcATTGT CTAATTGATA	900
15	TCCAAACTCA CTTAGACGTT GACCGCCCCA TAATTTTGTT TTTAAaATTG GTTGTAaaaa	960
	TAATGGCATT GTTGcACCTc CATTGTGATT AAGTAAGCAa TAGAACTCTG ATGTTGTtGT	1020
	TCCATTATAT TTTGATTTTG TTCTcATTTA CATCGTATTA TTAACtTCCA CATTTCAAAT	1080
20	TAActATTAG TGATTGTACC ATATTTACTA ACATTGCAGT ACTGCCaATT AAAAGhGCTT	1140
	CACTTAAaTT TACAGTACTT TAACATTTTc AAAAATTtAT AGCATAGAGA TTATATCTCT	1200
	CTTACATTTG TACATATTTc CCTTTAAaTT TACTCGCCCA TTATACCAAT TAATAaACAA	1260
25	CTTTAATAGT TGTGCCATAC ATTGTTCAAA TTCTTTGTAA AACGCATAGA CAATACGTAC	1320
	TTATTcATAC TTATAATTCA TCATTTTCAA AAAATAACGA GTTACGAAAA AGTAACCCGC	1380
	TTCAaATCAT ATTTACTATC CTTATTaATC CGTTTCATT TCAaATTGAG TTAAAGCATC	1440
30	TTTAATGTCC TGATCACCAC TAATAATTG AAACTCTTGG TGATTAAaAT GATTGGATGT	1500
	GACAATTTCT TTTAATACTG TCGCAACATC TTCTCTAGGA ATTTcACCTT TACCATCAAA	1560
35	ATATTGTGCA GCTTCTATCT TTCCAGATCC TGCTGCATT TTAAGTGCCC CTGGATGTAA	1620
	AATTGTATAA TTCAaACCTG hAACGTCTTA AATAGTCATC AGCGTAATGT TTAGCTATTG	1680
	TATATGGCTT TAAATCACCg CTATCATCAa AAGCCTGACG TCTCGAATCA TATGTTGAAA	1740
40	CCATGACATA GTGTTTAATA TTGGCCTCTT TACTCGCAAT CATTGATTTA ACAGCACCAT	1800
	CTAAATCGAC AATAATTGTT TTATCTGCAC CCGTGTTCCC TCCAGAACCT ACTGAAAAGA	1860
	TAActTTATC GAATGGTTTA AACGTCTCAG TTAAAGTCTC TATTGAATCA TTTTCAACAT	1920
45	CAACAAGAAT TGCTTTcATA CCTTGtGATT TTAACGCATT AAGTTGATCT GATTGCCTAA	1980
	CACCAGCAGT AAATGGTACA TTTTCTTTTG CTAATTGTTG CACTAGTAAC GAACCTACAC	2040
50	CGCCATTAGC ACCTATAACC AAAATATTCA TTTACAACAC TCTCCTATkT ATTATTCTCT	2100
	ATGCCATACC ACTTTATGAG ATATGTAAAA CTTGTTACAA CTATAAAaAT CAATTGACAT	2160
	ACTACTGGGA ACGTATTAAA TTAATATATG AACAAATATT CATATGAAAG GATTGTcATA	2220
55		

	tCaAGGCATT AGcGATTACA ATCGAATACG TATCaTGGAA TTGTTATCaG TCAGCGAAgC	2340
	AAGTGTGGT CACATTtCAC ATCAATTGAA TTTATCTCAA TCAAATGTCT CGCACCAATT	2400
5	AAAATTACTT AAAAGTGTGC ATCTTGTGAA AGCAAAACGA CAAGGCCAAT CAATGATTTA	2460
	TTCATTAGAT GACATCCACG TAGCAACTAT GTTAAAGCAA GCCATACATC ACGCGAATCA	2520
10	TCCTAAAGAA AGTGGGTTAT AATATGTCTC ATTACATCA TCATCATGAC CATATGCATA	2580
	GTCATGTAAC TACAAATAAT AAGAAAGTAT TGTTTATATC GTTTTAAATA ATCGGTCTAT	2640
	ATATGTTTAT CGAAATCATC GGCGGTCTCC TTGCTAACAG CTTGGCATTa CTATCTGACG	2700
15	GTATCCATAT GTTTAGCGAC ACATTCTCAT TAGGTGTTGC ACTTGTCGCA TTTATTTATG	2760
	CTGAAAAGAA TGCCACAACt ACAAAAACAT TTGGTTATAA ACGTTTCGAA GTACTCGCAG	2820
	CGTTATTTAA CGGTGTAACG CTTTTTGTAa TAAGTATTTT GATTGTTTTT GAAGCGATTA	2880
20	AACGTTTCTT TGTCCTTCT GAAGTTCAAT CAAAAGAAAT GTTAATCATT AGTATTATCG	2940
	GTTTAATGT CAATATCGTT GTTGCATTCT TTATGTTTAA AGGCGGCGAC ACTTCACACA	3000
	ATTTAAATAT GCGTGGTGCT TTTCTACATG TTATCGGAGA CTTATTAGGT TCAGTTGGCG	3060
25	CCATTACTGC AGCTAkTTTA ATTTGGGCAT TTGGATGGAC AATCGCCGAT CCTATCGCAA	3120
	GTATTTTAGT TTCCGTTATT ATTTTAAAAA GTGCTTGGGG TATCACAAAA TCTTCAATTA	3180
30	ACATTTTAAT GGaAGGCACA CCAAGTGATG TTGATATAGA TGAAGTTATA ACTACTATTA	3240
	AAAAGGATTC ACGAATACAA AGTGTGCATG ATTGCCATGT TTGGACAATT TCAAATGATA	3300
	TGAATGCATT GAGTTGTCAT GTTGTGTAG ACCATACATT GACAA TGAAa GAATGTGAAT	3360
35	TATTATTAGA AAaCATTGAG CATGATTTAT TACATTTAAA TATTCACCAT ATGACTATTC	3420
	AATTAGAAAC GCCTAATCAC AAACATGATG AATCGATTAT ATGTTcAGGA ACACATAGTC	3480
	ATTCaCATAA CCATCATGCT CATCATCAG CGCATGTACA TTAATAATTT TAACCTACTG	3540
40	CCATTGCATC GAITAAACTT TTCAATGGCA GTAGGTTTTT TATGTCTTTA TGGCGACTTG	3600
	TTTGGTCTTT GATGATGCAA TGTTTATTAA CAAATTTTCA ACTATTATTT CTTACATTAG	3660
	TCATATTTT GACAATTTAC TATTATAATT CTCTAACTTT AGTCACTTTA ATTAATTTTT	3720
45	ATTAGATATT AATATGAAAA TAACGTGTTT TTTGTTATT	3759

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13086 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear



EP 0 786 519 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

	TAATTATCGC GCATAACAAA ACATTAGCAG GACAAATTATA TAGTGAGTTT AAAGAATTTT	60
5	TTCCTGAAAA CAGGGTGGAA TACTTTGTAA GTtACTATGA TTATTATCAn CCAGAGGCAT	120
	ACGTACCGTC TACTGACACT TTTATTGAAA nAGATGCCTC AATCAnTGAT GAAATTGATC	180
	AACTACGACA TTCTGCTACA AGTGCAATTAT TTGAACGCGA TGATGTAATT ATTATTGCTA	240
10	GTGTAAGTTG TATATATGGT TTAGGTAATC CTGAAGAATA TAAAGATTTA GTAGTAAGTG	300
	TTCCAGTTGG TATGGAAATG GATAGAAGTG AATTACTTAG AAAACTTGTc AGATGTGCAA	360
	TATACACGAA ATGACATCgA TTTcCAACGA GGAACGTTTC GAGTGCGTGG TGATGTAGTG	420
15	GAAATATTCC CAGCCTCTAA AGAAGAACTT TGTATAAGGG TTGAGTTTTT CGGCGATGAG	480
	ATTGACCGTA TCCGAGAAGT TAACTACCTA ACAGGTGAAG TGTTGAAAGA AAGAGAACAT	540
20	TTTGCGATAT TCCCAGCTTC TCACTTCGTA ACACGTGAAG AAAAGTTGAA AGTTGCGATT	600
	GAACGTATTG AAAAAGAATT GGAAGAACGA TTGAAAGAAT TACGAGATGA GAATAAATTA	660
	CTAGAAGCGC AAAGGTTAGA ACAGCGTACC AACTATGATT TAGAAATGAT GCGAGAGATG	720
25	GGATTCTGTT CAGGAATTGA AAATATTCC GTACATTAA CTTTGCGACC ACTGGGTTCC	780
	ACACCATATA CTTTATTGGA TTACTTTGGC GATGATTGGT TAGTAATGAT TGATGAATCA	840
	CATGTGACAT TACCGCAAGT TCGAGGCATG TATAACGGAG ACAGAGCGCG TAAACAAGTT	900
30	TTGGTGGATC ATGGGTTTAG ATTACCGAGT GCATTAGATA ACCGTCCACT TAAATTGAA	960
	GAATTTGAAG mAAAGACAAA ACAACTTGTG TATGTATCTG CAACGCCTGG ACCATACGAA	1020
	ATTGAACATA CGGATAAGAT GGTGGAACAA ATTATTGCTC CTACTGGTTT ACTGGATCCT	1080
35	AAGATTGAGG TTAGACCTAC TGAAAATCAA ATTGACGATT TATTAAGTGA AATTCAAACA	1140
	AGAGTgAGCG TAATGAACGC GTACTTGTTA CAACGCTCAC TAAAAAGATG AGTGAAGATT	1200
40	aACCACATAC ATGAAAGaG CGGGTATTAA aGtTAATTAT CTGCATTcAG AAATCAAGAC	1260
	ATTAGAACGA ATTGAAATAA TTAGAGACTT ACGAATGGGT ACATATGATG TTATCGTAGG	1320
	TATTAATTTA TTAAGAGAGG GTATTGATAT ACCAGAAGTT TCTCTAGTTG TCATATTAGA	1380
45	TGCAGATAAA GAAGGGTTTT TACGTTCTAA CCGCTCATTa ATTCAAaCAA TAGGTAGAgC	1440
	TGCGCGTAAC GATAAaGGTG AAGTCATTAT GTATGCCGAT AAAATGACTG ATTcGATGAA	1500
	GTATGCAATT GATGAGACAC AACGTCGTCG AGAAATACAG ATGAAACATA ATGAAAAACA	1560
50	TGGTATTACA CCTAAAACAA TTAATAAAAA AATACATGAT TTAATTAGTG CTACTGTGTA	1620
	AAATGACGAA AATAATGACA AAGCACAaAC TGTGATACCT AAGAAGATGA CGAAAAAAGA	1680

55

	TTTCGAGAAA GCTACAGAAT TAAGAGATAT GTTATTTGAA TTAAAAGCAG AAGGGTGACA	1800
	AGTAAATGAA AGAACCATCC ATAGTAGTAA AAGGTGCTCG TGCGCATAAC TTGAAAGATA	1860
5	TTGATATCGA ACTACCTAAA AaTAAATTAA TTGTTATGAC AGGTTTATCT GGGTCAGGTA	1920
	AATCGTCATT AGCATTTCGAT ACTATATATG CTGAAGGACA ACGACGTTAT GTTGAATCAT	1980
	TAAGTGCCTA TGCGCGTCAA TTTTtagGCC AAATGGACAA ACCAGATGTT GATACAATTG	2040
10	AAGGATTATC GCCAGCAATT TCAATAGATC AAAAAACAAC AAGTAAAAAT CCAAGATCAA	2100
	CTGTAGCAAC AGTAACAGAA ATATATGATT ATATACGTTT GTTATATGCA CGTGTGTA	2160
15	AACCTTACTG TCCAAATCAC AATATAGAAA TTGAATCGCA AACAGTACAA CAAATGGTTG	2220
	ACCGCATTAT GGAATTAGAG GCACGTACAA AGATTCAATT ATTAGCACCT GTCATCGCTC	2280
	ATCGTAAAGG TAGTCATGAA AAGCTAATCG AAGATATTGG TAAAAAAGGT TATGTACGTT	2340
20	TAAGAATCGA TGGCGAAATT GTTGATGTAA ATGATGTACC TACTTTAGAT AAGAACAAGA	2400
	ATCATACAAT AGAAGTTGTT GTAGACCGAT TAGTTGTTAA AGATGGAATT GAAACACGAC	2460
	TAGCTGACTC TATAGAAACT GCCTTAGAGC TTTCAGAAGG ACAATTAACA GTCGATGTCA	2520
25	TTGACGGGGA AGACCTTAAG TTTTCAGAAA GCCATGCTTG TCCTATATGT GGATTTTCAA	2580
	TCGGAGAGTT AGAACCAAGA ATGTTTAGCT TTAACAGTCC TTTTGGTGCT TGTCCGACAT	2640
	GTGATGGCTT AGGCCAAAAG TTAACAGTCG ATGTAGACTT GGTGTGTTCC GACAAAGATA	2700
30	AGACGCTAAA CGAAGGTGCA ATAGAACCTT GGATACCGAC GAGTTCTGAT TTTTATCCAA	2760
	CATTGTTHAA ACGTGTGTTGT GAAGTTTATA AAATCAATAT GGATAAACCT TTTAAAAAGT	2820
	TAACAGAACG TCAACGTGAT ATTTTATTGT ATGGTTCTGG TGACAAAGAA ATTGAATTTA	2880
35	CATTTACACA ACGTCAAGGT GGTACTAGAA AACGAACAAT GGTTTTCGAG GGTGTAGTTC	2940
	CTAAATATAAG TAGACGATTC CATGAATCTC CTTCAGAATA TACACGTGAA ATGATGAGTA	3000
40	AATATATGAC TGAACCTACCT TGCGAAACTT GTCATGGAAA GCGATTGAGT CGTGAAGCKT	3060
	TATCTGTTTA TGTAGGTGGT TTAAATATTG GTGAAGTAGT CGAATATTCA ATCAGTCAAG	3120
	CGCTGAAC TAATAAAAAC ATTGATTTGT CAGAACAAGA TCAAGCGATT GCAAATCAAA	3180
45	TATTGAAAGA AATTATTTCC CGACTCACTT TTTTAAATAA TGTGGGACTT GAATATTTAA	3240
	CGTTAAACAG AGCTTCAGGT ACACCTTCAG GTGGTGAAGC ACAACGTATT CGATTAGCAA	3300
	CGCAAATTGG GTCGCGTTTG ACTGGTGTCT TATATGTATT AGATGAGCCA TCAATTGGAC	3360
50	TGCATCAAAG AGATAATGAT CGATTAATTA ATACACTTAA AGAAATGAGA GATTTAGGAA	3420
	ATACTTTAAT TGTAGTTGAA CACGATGATG ATACAATGCG TGCGGCTGAT TACTTAGTGG	3480

55

EP 0 786 519 A2

	AGGTAATGAA AGATAAAAAA TCATTAAACAG GACAATACTT GAGTGGTAAG AAACGTATTG	3600
	AAGTACCTGA ATATCGCAGA CCGGCTTCAG ATCGTAAAT TTCTATACGT GGAGCTAGAA	3660
5	GCAACAATCT TAAAGGGGTT GATGTGGACA TACCACTATC AATCATGACG GTTGTTACAG	3720
	GTGTATCAGG TTCTGGTAAA AGCTCATTAG TAAATGAAGT ATTATACAAA TCATTAGCTC	3780
	AAAAAATTAA TAAATCTAAA GTAAAGCCAG GATTGTACGA TAAGATTGAA GGTATTGATC	3840
10	AACTTGATAA AATTATTGAT ATTGATCAAT CACCAATAGG TAGAACGCCA CGCTCTAATC	3900
	CAGCAACATA TACTGGTGTG TTTGATGATA TACGTGATGT GTTTGCGCAA ACAAATGAAG	3960
	CTAAAAATTCG AGGATATCAA AAAGGGCGTT TTAGTTTAA TGTAAGAGGT GGACGCTGTG	4020
15	AAgcTTGTAA AGGTGACGGT ATTATTAAAA TTGAAATGCA TTTTTTACCT GATGTTTATG	4080
	TTCTTGTTGA AGTGTGTGAT GGTAAACGAT ATAATCGTGA GACACTAGAG GTTACTTACA	4140
20	AAGGTAAAAA TATTGCTGAC ATTTTAGAAA TGAAGCAACA CAATTTTGTG	4200
	AAAATATTCC TAAGATTAAG CGCAAGTTAC AAACACTAGT TGATGTTGGT CTTGGATACG	4260
	TCACATTAGG TCAACAAGCT ACAACGTTAT CAGGTGGTGA GGCTCAACGT GTGAaACTTG	4320
25	CATCTGAAC TCAATAACGT TCAACTGGTA AATCTATTTA TATCCTAGAT GAACCGACAA	4380
	CAGGGTTACA TGTGACGAT ATTAGTAGAT TATTAAGAGT ATTAAACCGA TTAGTTGAAA	4440
	ATGGTGATAC TGTGTAATT ATTGAACATA ACCTAGATGT TATCAAAACA GCAGACTATA	4500
30	TTATAGACTT AGGTCCTGAA GGTGGTAGTG GCGGTGGTAC TATTGTTGCG ACTGGCACAC	4560
	CCGAAGATAT TGCTCAGACA AAGTCATCAT ATACAGGAAA GTATTTAAAA GAAGTACTTG	4620
	AACGAGATAA ACAAATACT GAAGATAAAT AAGATTAAAA GAAGTGAAGG ATGTTATAAA	4680
35	TTTATCCTTC GCTTCTTTTT ATTAATTTAG TAATGAATAG TAGAAAGAAA AGATGCGTAA	4740
	AAAGaATTAT GTTAAGATAG GGTCAATCTA GAGTAGTTAA ACATAAATCG AACTGGGAGT	4800
40	GGGACAGAAA TGATAAAGAA TCACTAATGA TTTATTATGT AGTGGTTCTT TGTCATTAGC	4860
	CACAGCTATT GTGTACTTAA AAATAGGaat GCaTgAGTGC AACTCATGCA TAAGaAATAC	4920
	TAATTTCTAA AGAAAAAGTA TTTCTTTATG TTGGGGCCCC GCCAACTTGC ATTGTTTGTA	4980
45	GAATTTCTTT TCGAAATCTT TTATGTTGGG GCCCCGCCAA CTTGCATTGT TTGTAGAATT	5040
	TCTTTTCGAA ATTCTTTATG TTGGGGCCCC GCCAACTAAT TCCAATATAT CATTGTAGAG	5100
	CTTAGGTCAT TGATTTTTGG CTCGGACTTT TATGGCGATA TGAACCATGT AAATTAAGCA	5160
50	AGCAATAAAT TAATGATTGA TATTGACTTG TAAAATAATA ACAATAATGA ACAATTAATA	5220
	TTTATTTTAG CTTTCAATG TAGATTGGTG TTATATTTTT GATATGATAA GAAGAGATGT	5280

55

## EP 0 786 519 A2

	ACATTAAAGT TAGATTTAAT CGCTGGTGAA GAAGGACTAT CGAAGCCAAT TAAAAATGCT	5400
	GATATATCAA GACCGGGCTT AGAGATGGCA GGTATTTTTT CACATTATGC GTCAGATAGA	5460
5	ATACAACTAT TAGGAACAAC GGAACATATCG TTTTACAATT TATTACCAGA TAAGGATCGC	5520
	GCAGGTCGTA TGCCTAAACT ATGCAGACCA GAAACGCCTG CAATTATTGT GACACGTGGA	5580
	TTGCAGCCAC CAGAAGAATT AGTTGAAGCT GCAAAAGAAT TAAATACCCC ACTTATAGTT	5640
10	GCTAAAGATG CCACTACAAG TTTAATGAGT CGCTTAACAA CGTTTTTAGA GCATGCACTT	5700
	GCAAAGACGA CATCTTTACA TGGTGTTTTA GTAGATGTTT ACGGTGTTGG TGTACTAATT	5760
	ACCGGTGATT CAGGAATAGG TAAAAGTGAG ACTGCGTTGG AATTAGTTAA ACGTGGGCAT	5820
15	AGATTAGTAG CAGATGATAA TGTAAGAAATA CGTCAAATTA ATAAAGATGA ACTAATAGGG	5880
	AAACCACCAA AGTTAATAGA ACATCTATTA GAAATACGTG GACTAGGTAT TATCAATGTT	5940
20	ATGACTTTAT TTGGCGCGGG TTCAATATTA ACTGAAAAC GAATTAGATT AAATATTAAT	6000
	TTGGAAGAACT GGAACAAGCA AAAGTTATAT GACCGCGTAG GTCTTAATGA AGAGACGCTA	6060
	AGTATTTTAG ATACTGAAAT CACTAAAAAA ACAATACCTG TAAGACCTGG TAGAAATGTT	6120
25	GCGGTAATTA TTGAGGTCGC TGCAATGAAC TATCGATTAA ATATCATGGG CATTACACG	6180
	GCCGAAGAAT TTAGTGAAAG ATTAAATGAA GAAATTATCA AGAACAGTCA TAAGAGTGAG	6240
	GAGTAGGTTG AATGGGTATT GTATTTAACT ATATAGATCC TGTGGCATT TAACTTAGGAC	6300
30	CACTGAGTGT ACGATGGTAT GGAATTATCA TTGCTGTCGG AATATTACTT GGTACTTTG	6360
	TTgCACAAACG TGCACTAGTT AAAGCAGGAT TACATAAAGA TACTTTAGTA GATATTATTT	6420
	TTTATAGTGC ACTATTTGGA TTTATCGCGG CACGAATCTA TTTTGTGATT TTCCAATGGC	6480
35	CATATTACGC GGAAAATCCA AGTGAAATTA TTAAAATATG GCATGGTGGA ATAGCAATAC	6540
	ATGGTGGTTT AATAGGTGGC TTTATTGCTG GTGTTATTGT ATGTAAAGTG AAAAAATTAA	6600
40	ACCCATTTC AATTGGTGAT ATCGTTGCGC CAAGTATAAT TTTAGCGCAA GGAATTGGAC	6660
	GCTGGGGTAA CTTTATGAAT CACGAGGCAC ATGGTGGATC GGTGTCACGC GCTTTTTTAG	6720
	AACAATTACA TTTGCCTAAT TTTATAATAG AAAATATGTA TATTAACGGC CAATATTATC	6780
45	ATCCAACATT CTTATATGAA TCCATTTGGG ATGTCGCTGG ATTTATTATC TTAGTTAATA	6840
	TTCGTAAACA TTTAAAATTA GGAGAAACAT TCTTTTTATA TTTAACTTGG TATTCAATTG	6900
	GTCGATTCTT TATAGAAGGA TTACGTACAG ATAGCTTAAT GCTCACAAGT AATATTAGAG	6960
50	TTGCACAATT AGTATCAATT CTTTTAATTT TAATAAGTAT AAGTTTAATT GTATATAGAA	7020
	GGATTAAGTA TAATCCACCG TTGTATAGCA AAGTTGGGGC GCTTCCATGG CCAACAAAAA	7080

55

## EP 0 786 519 A2

	TTATGGCGTG TATACCGTCT TGTAAATTT TCGAAAGTTT TTAAGAATGT AATTATCATT	7200
	GAATTTTCGA AATTTATTCC AAGTATGGTA CTGAAAAGAC ATATATATAA ACAACTTTTA	7260
5	AATATTAATA TCGGTAATCA ATCGTCGATA GCTTATAAAG TAATGTTAGA TATTTTTTAC	7320
	CCAGAACTGA TTACGATTGG TAGTAACAGT GTTATTGGTT ACAATGTAAC AATTTTGACG	7380
	CATGAAGCAT TAGTTGATGA ATTTTCGTAT GGACCAGTGA CGATAGGATC TAACACTTTG	7440
10	ATTGGTGCAA ATGCTACCAT TTTACCCGGT ATAACGATTG GTGACAATGT AAAAGTTGCA	7500
	GCTGGTACGG TTGTTTCAAA AGATATACCG GATAATGGAT TTGCATATGG CAACCCTATG	7560
	TATATAAAAA TGATTAGGAG GTGACAATTT TATGGCGCAA AAGAATAATA ATGTAATTCC	7620
15	AATGACTTTT GATGATGCAT TTTATCGTAA AATGGCTAAA CAGAAGTTTA AACAAAGAGA	7680
	ATATAAACGA GCTGCTGAAT ACTTTGAAAA AGTGTTAGAA TTGTCACCTG ATGATCTGGA	7740
20	AATTCAAATT GATTATGCAC AATGTCTAGT GCAACTTGGT ATTGCTAAAA AAGCAGAACA	7800
	TTTATTTTAT GACAATATTA TTTATAATAG GCATCTAGAA GATAGCTTTT ATGAATTGAG	7860
	TCAGCTCAAC ATTGAAGTTA ACGAACCAAA CAAGGCATTG TTGTTTGGA TTAATTATGT	7920
25	TATTGTTAGC GACGACCAAG ATTATAGAGA TGAATTAGAT CAAATGTTTG ATGTGAAATA	7980
	TCAAAGTGAA GAACAAATTG AACTTGAAGC TCAATTGTTT GTAGTTCAA TACTATTCCA	8040
	ATATCTTTTT TCTCAAGGTC GATTAAAAGA TGCAAAGAAT TATGTCTTAC ATCAACCACA	8100
30	AGAAGTTCAA GATCATCGTG TAGTACGTAA TTTATTGGCA ATGTGTTATT TATATCTCGG	8160
	TGAATATGAT ACgGCTAAAG CATTGTACGA aGCACtATTA CAAGAGGATA GTACaGATAT	8220
	ATATGCATTA TGCCATTATA CTTTGCTACT TTATAACACT AAGGAAAATG AACAATATCA	8280
35	AAAATATTTA AAAATATTAA ACAAAGTTGT ACCTATGAAT GACGATGAAA GTTTTAAATT	8340
	AGGTATTGTA TTAAGTTATT TAAAGCAGTA TCGTGCATCA CAACAATTGT TGTACCCTTT	8400
40	ATATAAAAAA GGGAAATTTT TATCAATTCA AATGTACAAT GCTTTAGCAT ATAATTATTA	8460
	TTATTTAGGT GAAGAAGACG AAAGTCATTA CTA CTGGGAT AAATTGAAGC AAATTTCTAA	8520
	AGTGGAATTT GGACATGCCG CTTGGGTAAT TGAAAATAGC AAAGAAGTTT TTGACCAACA	8580
45	TATTTTGCCA TTA CTTCAAA GTGATGACAG TCATTATCGT TTATATGGTA TTTTTTTATT	8640
	GGATCAATTA AATGGTAAAG AAATTGTGAT GACGGAAAGT ATTTGGCAGG TTTTGAAAA	8700
	TCTAAATAAT TATGAGAAAT TGTATTTAAC GTATTTAGTT CAAGGTTTAA CGCTCAATAA	8760
50	ATTAGACTTC ATTCATCGCG GCTTATTAAC GCTTTACCAT AATGAATTAT TTGTAAGTGA	8820
	AAATGATGTA ATGGTTGCAT GGATTAATCA AGGTGAAGTC ATAATTGCTG AAAAAGTAGA	8880
55		

## EP 0 786 519 A2

	TCGAAACGTT	ACAAAGAAGC	AAATTACAAC	ATGGTTAGGC	ATAACACAAT	ATAAACTGAA	9000
	CAAAATGATT	GAATTTCTCT	TGAGCATATA	GATTTATGAA	AAGTTAGATT	TATTATATAA	9060
5	TGCGCATAAT	GATTAATAAT	GAGGAGGCGT	TAATAAAATG	ACTGAAATAG	ATTTTGATAT	9120
	AGCAATTATC	GGTGCAGGTC	CAGCTGGTAT	GACTGCTGCA	GTATACGCAT	CACGTGCTAA	9180
	TTTAAAAACA	GTTATGATTG	AAAGAGGTAT	TCCAGGCGGT	CAAATGGCTA	ATACAGAAGA	9240
10	AGTAGAGAAC	TTCCCTGGTT	TCGAAATGAT	TACAGGTCCA	GATTTATCTA	CAAAAATGTT	9300
	TGAACACGCT	AAAAAGTTTG	GTGCAGTTTA	TCAATATGGA	GATATTAAAT	CTGTAGAAGA	9360
	TAAAGGCGAA	TATAAAGTGA	TTAACTTTGG	TAATAAAGAA	TTAACAGCGA	AAGCGGTAT	9420
15	TATTGCTACA	GGTGCAGAAT	ACAAGAAAAT	TGGTGTTCGG	GGTGAACAAG	AAC TTGGTGG	9480
	ACGCGGTGTA	AGTTATTGTG	CAGTATGTGA	TGGTGCATTC	TTTAAAAATA	AACGCCTATT	9540
20	CGTTATCGGT	GGTGGTGATT	CAGCAGTAGA	AGAGGGAACA	TTCTTAACTA	AATTTGCTGA	9600
	CAAAGTAACA	ATCGTTTACC	GTCGTGATGA	GTTACGTGCA	CAGCGTATTT	TACAAGATAG	9660
	AGCATTCAAA	AATGATAAAA	TCGACTTTAT	TTGGAGTCAT	ACTTTGAAAT	CAATTAATGA	9720
25	AAAAGACGGC	AAAGTGGGTT	CTGTGACATT	AACGTCTACA	AAAGATGGTT	CAGAAGAAAC	9780
	ACACGAGGCT	GATGGTGTAT	TCATCTATAT	TGGTATGAAA	CCATTAACAG	CGCCATTTAA	9840
	AGACTTAGGT	ATTACAAATG	ATGTTGGTTA	TATTGTAACA	AAAGATGATA	TGACAACATC	9900
30	AGTACCAGGT	ATTTTTGCAG	CAGGAGATGT	TCGCGACAAA	GGTTTACGCC	AAATTGTCAC	9960
	TGCTACTGGC	GATGGTAGTA	TTGCAGCGCA	AAGTGCAGCG	GAATATATTG	AACATTTAAA	10020
	CGATCAAGCT	TAATTCGAAG	TCGAATTAAG	ATGTTGAGCT	GTAAATTATT	TGGATATTTA	10080
35	TTTTAATAGT	GTCATCACAG	CGTTAAAATA	ATGTCTTACT	TTTAAATTAA	AGCAAATTAT	10140
	ATAGAAAAC	AGAACTTAGT	ACGTATCATT	TGTGCGTTTC	AATGAGTTCT	AGTTTTTTTTA	10200
	TATGTTATAT	TAAACTTATA	ACTTTATGGG	AGTGGGACAG	AAATGATAAA	GAGCCACTAA	10260
40	TGATTTATTA	TGTAGTGGTT	CTTAAACATT	AGCCACAGCT	AATGTGTACT	TAAAAATAGG	10320
	AATACATGAG	TAAAACTCAT	GCATAAGAAA	TACTAATTTT	TATAGAAAAA	GTATTACTTT	10380
45	ATCGTTGTCC	CACCCCAACT	TGCACATTAT	TGTAAGCTGA	CTTCCGCCA	GCTTCTGTGT	10440
	TGGGGCCCCG	CCAAC TTGCA	CATTATTGTA	AGCTGACTTT	TCGTCAGCTT	CTGTGTTGGG	10500
	GCCCCGCCAA	CTTGACATT	ATTGTAAGCT	GACTTTTCGT	CAGCTTCTGT	GTTGGGGCCC	10560
50	CGCCAAC TTG	CATTGTCTGT	AGAAATTGGG	AATCCAATTT	CTCTATGTTG	GGGCCCACAC	10620
	CCCAACTCGC	ATTGCCTGTA	GAATTTCTTT	TCGAAATTCT	CTGTGTTGGG	GCCCACACCC	10680

55

EP 0 786 519 A2

	ACTCGCATTG	CCTGTAGAAT	TTCTTTTCGA	AATTCTCTGT	GTTGGGGCCC	CTGACTAGAG	10800
	TTGAAAAAAG	CTTGTTGCAA	GCGCATTTC	ATTCAGTCAA	CTACTAGCAA	TATAATATTA	10860
5	TAGACCCTAG	GACATTGATT	TATGTCCCAA	GTCCTTTTA	AATGATGTAT	ATTTTITAGAA	10920
	ATTTAATCTA	GACATAGTTG	GAAATAAATA	TAAAACATCG	TTGCTTAATT	TTGTCATAGA	10980
	ACATTTAAAT	TAACATCATG	AAATTCGTTT	TGGCGGTGAA	AAAATAATGG	ATAATAATGA	11040
10	AAAAGAAAAA	AGTAAAAGTG	AACTATTAGT	TGTAACAGGT	TTATCTGGCG	CAGGTAAATC	11100
	TTTGTTTATT	CAATGTTTAG	AAGACATGGG	ATATTTTTGT	GTAGATAATC	TACCACCAGT	11160
	GTTATTGCCT	AAATTTGTAG	AGTTGATGGA	ACAAGGAAAT	CCATCCTTAA	GAAAAGTGGC	11220
15	AATTGCAATT	GATTTAAGAG	GTAAGGAACT	ATTTAATTCA	TTAGTTGCAG	TAGTGGATAA	11280
	AGTCAAAAGT	GAAAGTGACG	TCATCATTGA	TGTTATGTTT	TTAGAAGCAA	GTAAGTAAAA	11340
20	ATTAATTTCA	AGATATAAGG	AAACGCGTCG	TGCACATCCT	TTGATGGAAC	AAGGTAAAAG	11400
	ATCGTTAATC	AATGCAATTA	ATGATGAGCG	AGAGCATTTG	TCTCAAATTA	GAAGTATAGC	11460
	TAATTTTGTT	ATAGATACTA	CAAAGTTATC	ACCTAAAGAA	TTAAAAGAAC	GCATTCGTCG	11520
25	ATACTATGAA	GATGAAGAGT	TTGAACTTTT	TACAATTAAT	GTCACAAGTT	TCGGTTTTAA	11580
	ACATGGGATT	CAGATGGATG	CAGATTTAGT	ATTTGATGTA	CGATTTTITAC	CAAATCCATA	11640
	TTATGTAGTA	GATTTAAGAC	CTTTAACAGG	ATTAGATAAA	GACGTTTATA	ATTATGTTAT	11700
30	GAAATGGAAA	GAGACGGAGA	TTTTCTTTGA	AAAATTAAGT	GATTTGTTAG	ATTTTATGAT	11760
	ACCCGGGTAT	AAAAAGAAG	GGAAATCTCA	ATTAGTAATT	GCCATCGGTT	GTACGGGTGG	11820
	ACAACATCGA	TCTGTAGCAT	TAGCAGAACG	ACTAGGTAAT	TATCTAAATG	AAGTATTTGA	11880
35	ATATAATGTT	TATGTGCATC	ATAGGGACGC	ACATATTGAA	AGTGGCGAGA	AAAAATGAGA	11940
	CAAAATAAAG	TTGTACTTAT	CGGTGGTGGC	ACTGGCTTAT	CAGTTATGGC	TAGGGGATTA	12000
40	AGAGAATTCC	CAATTGATAT	TACGGCGATT	GTAACAGTTG	CTGATAATGG	TGGGAGTACA	12060
	GGGAAAATCa	GAGATGAAAT	GGATATACCA	GCACCAGGAG	ACATCAGAAA	TGTGATTGCA	12120
	GCTTTAAGTG	ATTCTGAGTC	AGTTTAAAGC	CAACTTTTTT	AGTATCGCTT	TGAAGAAAAT	12180
45	CAAATTAGCG	GTCATCATT	AGGTAATTTA	TTAATCGCAG	GTATGACTAA	TATTACGAAT	12240
	GATTTTCGGAC	ATGCCATTAA	AGCATTAAGT	AAAATTTTAA	ATATTAAAGG	TAGAGTCATT	12300
	CCATCTACAA	ATACAAGTGT	GCAATTAAAT	GCTGTTATGG	AAGATGGAGA	AATTGTTTTT	12360
50	GGAGAAACAA	ATATTCCTAA	AAAACATAAA	AAAATTGATC	GTGTGTTTTT	AGAACCTAAC	12420
	GATGTGCAAC	CAATGGAAGA	AGCAATCGAT	GCTTTAAGGG	AAGCAGATTT	AATCGTTCTT	12480
55							

GCGTTAATTC ATTCTGATGC GCCTAAGCTA TATGTTTCTA ATGTGATGAC GCAACCTGGG 12600  
 GAAACAGATG GTTATAGCGT GAAAGATyAT ATCGATGCGA TTCATAGACA AGCTGGACAA 12660  
 5 CCGTTTATTG ATTATGTCAT TTGTAGTACA CAAACTTTCA ATGCTCAAGT TTTGAAAAAA 12720  
 TATGAAGAAA AACATTCTAA ACCAGTTGAA GTTAATAAGG CTGAACTTGA AAAAGAAAGC 12780  
 ATAAATGTAA AAACATCTTC AAATTTAGTT GAAATTTCTG AAAATCATTT AGTAAGACAT 12840  
 10 AATACTAAAG TGTTCATGAC AATGATTTAT GACATAGCTT TAGAATTAAT TAGTACTATT 12900  
 CCTTTCGTAC CAAGTGATAA ACGTAAATAA TATAGAACGT AATCATATTA TGATATGATA 12960  
 ATAGAGCTGT GAAAAAATG AATATAGACA GTGGTTCTAA GGTGAATCAT GTTTTAAATA 13020  
 15 AGAAAGGAAT GACTGTACGA TGAGCTTTGC ATCAGAAATG AAAAATGAAT TAACTAGAAT 13080  
 AGACGT 13086

20 (2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1350 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

30 CATTAGTCAT GAAAATAGCC GACAACTTCA TCTGTGAAAT CACCGGCCTT TTATTTTAGC 60  
 TAACTTTATT TCTGATTTTA CGATTTTAAT TGATCATACA GAGAAAGTGA TCTTTTACAA 120  
 ATTTCTAAAA ACTCATGATC TATATTGGAC ATTTGATGAA AATAAGACAA AATGTTTTCT 180  
 35 GTTAGCTTCT CTTGTTTTGG GAATGAATCA TCTTCTTTAA TCCAAATCGC TAATTCGCCT 240  
 AATGTTGTTT TATCATCTTT AAATGTTTGT ATATATTCGT AAAAGCTCAT AGTATTCCTT 300  
 CTCTCAATT ACTTATATAA ATCCTACCAC GAAAGCTTTC AAGAAAACAC AATTAAATGT 360  
 40 CTATTTAGTG AACTTTTTAA GGTGTGTCAC TCTTTTAATG TCTGCCAATT AGGTCAATTA 420  
 ATCATCACAA TGTACAATTA ACTCTATTTT CAGTTCATAT ACTCACACAC CGTTTTTGAA 480  
 45 CAACACATTA ACTTCTCATT TAGATAAAAC GCAAAAAGC CTGGCACCAA TACAATAGAT 540  
 GCCAGACTAA GAGTCTACTA TATAAATTTA TTAGCGTAT GGTTTTACTT CGATTGCACC 600  
 TTCATTTTCA TCATGAACAC CATGCTTATA ATAATCAATA TATTGTGGCT CTAAAGGCTT 660  
 50 TCTGCCACGT ATAATGTCTG CTGCTTTTTT AGCTAACATT AAAACAGGTG CGTGTATATT 720  
 GCCATTTGTC GTACGTGGCA TAGCTGATGC ATCAACTACA CGTAAATTTT CCATACCGTG 780

55



ACTACAAGAT GGGTGTAAATG CTGTTTCACC ATCTCTACGA ACCCAATCAA GAATTTCTTC 900  
 GTCTGTTTGC ACTTCTGGTC CTGGTGAAAT TTCTCCACCA TTGAATGGAT CCATTGCTTT 960  
 5 TTGAGATAAG ATATTTCTTG CTACACGAAT TGCTTCTACC CATTCTTTTT TATCTTCTTC 1020  
 TGTGATAAA TAATTAAAGC GGATACTTGG TTTTTCGAAT GGATCTTTAG ATTTGATTTT 1080  
 CAAGCTACCA CGAGAGTTTG AATACATTGG TCCTACGTGA ACTTGATAAC CATGTGCGAC 1140  
 10 CGCTGCCTTT TGACCATCAT ATCTTACAGC TATTGGTAAG AAATGGAACA TTAAGTTAGG 1200  
 ATAAtCAACT TCGTTATTTG AACGTACAAA TCCGCCACCT TCAAAATGGT TAGATGCTGC 1260  
 15 TGCACCTGTA CGTGTGAAAA TCCATTGTAA ACCAATAAAT GGcATGCGCT TGAtATCTAA 1320  
 GCTTGGCtGt AATGATACAG GTTCCTTACA 1350

## (2) INFORMATION FOR SEQ ID NO: 18:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1376 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

TAATGCTATT GGCAACACCA TATATGAAAn CTCCAAACGA TCCTAAACCG ACTATAGATT 60  
 30 CACCAAATTT nACAATCCAT GAATAAGTA GTGGCCATAA GAATAACAAT ATGACAACCTA 120  
 AAAATGTACA GTAAATGCA GTCATAATTG GAACTAGACG TTTACCACTA AAAAATGATA 180  
 ATGCTAATGG TAATTCTGTT TCACTAAACT TATTGTATGC ATAAGCTGCT ATTAAACCTA 240  
 35 TTACAATACC AACAAAGACA TTGCCATTAT TCATCTTTTC AAAAGCTGAA TTTATTTCCTG 300  
 ArGCTTTCAT TCCTAATAAA GCGGCTAATT TCATTGGTGA TAATACAACT GTAACATAAA 360  
 40 AATATCCTAA CGTrGCTGCA rCGGsGACTG CACCATCATT TTTCTTTGCC ATTCCTATAG 420  
 CTACACCAAT TGCAAATAAA ATACCTAATT GCTCTAAAAT CGTAGTACCT ACCGTAGTAA 480  
 AGAACATTGC GATTTTCGGC GTCGCATGAA GTGCATTTAA CGTATTACCA ATTCCGGCAA 540  
 45 TAATTGCTGC AGCCGGTAAA ATGGCAACTG GTAACATTAA CGAACGCCCT AAATTTTGGA 600  
 AAAATTTATA CATTGAATGT CATCCTTCTT AAAATAATGT AGAAATATAA AGATTACTAA 660  
 TGTAAC TAGA ATAAC TACTT CGATACTCCG TTATAGTCAC CTAGGCTTAC TAACCAGCTA 720  
 50 TATTTCTACC TCAAGTTATT TTATAAACTT TTTACAATTT CATGCAATTC TTGTTGTAAC 780  
 TTTGCTGTTT GTGTTTCAAT CTCTTTTGTA ATATAATCGA TACGCTCGTT TCGTTTTAAA 840

55

AAAGACCGTG AATCTTAGTA GGACCAACAT AAGCAACAGG TAATATTGGT GACTTACTTA 960  
 ACATTGCAAT TGTGAAGCA CCaCGTTTCA AAGGTGCACC TTCTTGCGAT GTGCGAGAAC 1020  
 5 CTGTTGGGAA GATACCAACT GTCTTATTAT CTTTCAACAA ATTGATTGGG CGTTTTAAAG 1080  
 TACTAGGTCC TGGATTTTCA CGATCTACAG GAAATGCATT TAAAGACGTT AAAAATTTAC 1140  
 CAATCCATT ATTTTTGAAT AATTCTTTTT TAGCCATATA ATGAATTTGA TTAGGATATA 1200  
 10 ATGCCATACC TAGCATAATG ACTTCGTTAT AACTTTCATG CGTACAAGTT ACGACATATT 1260  
 TACTATCCTT AGGAATATTA TCTTTACCGA TTACGTATAA TGATTTTGAC ATTTTAACTA 1320  
 AAATGAAATT CAAAATCTTA CTAATCACTG AATACATTGT GCCACCTACT TAACTT 1376  
 15

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7363 base pairs  
 (B) TYPE: nucleic acid  
 20 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TTGTCATACC AATATTTTGT AAAATATGGA ACACAAGTAA AGTGACGAAA CCAACGATAA 60  
 AGATTTTGTT AAATTGATCT TCAATTTTCG CAGCTAATCT TATTAGATGG AAGATTAAAA 120  
 30 ATAAAAATAT TAAGATCAAT ATGACAGAAC CGATAAAGCC AAGTTCCTCT CCAATCACTG 180  
 AAAAGATAAA GTCAGTATGA TTTTCAGGTA TATAAACTTC ACCGTGATTG TATCCTTTAC 240  
 CTAGTAACTG TCCAGAACCG ATAGCTTTAA GTGATTCACT TAAATGaTAG CCATCACCAC 300  
 35 TACTATATGT ATAGGGGTCA AGCCATGAAT TGATTCGTCC CATTGATAC AGTTGGaCAC 360  
 CTAATJAAAT TTCAATTAAT GCGGGTGCAT ATAGaATACC TAAATGACT GTCATTGCAC 420  
 CAACaATACC TGTAATAAAG ATAGGTGCTA AGATACGCCA TGTTATACCA CTTACTAACA 480  
 40 TCACACCTGC AATAATAGCA GCTAATACTA ATGTAGTTCC TAGGTCATTT TGCAGTAATA 540  
 TTAAAATACT TGGTACTAAC GAGACACCAA TAATTTTGAA AAATAATAAC AAATCACTTT 600  
 GGAATGATTT ATTGAATGTG AATTGATTAT GTCTAGAAAC GACACGCGCT AATGCTAAAA 660  
 45 TTAAAATAAT TTTCATGAAT TCAGATGGCT GAATACTGAT AGGGCCAAAC GTGTACCAAC 720  
 TTTTGGCACC ATTGATAATA GGTGTAATAG GTGACTCAGG AATAACGAGC AAGCCTATTA 780  
 50 ATAATAGACA GATTAAAGAA TACAATAAAT ATGTATAATG TTTAATCTTT TTAGGTGAAA 840  
 TAAACATGAT GATACCTGCA AAAATTGCAC CTAAAATGTA ATAAAAAATT TGTCTGATAC 900

55

	TTGCTAAAAC AGCTATAGTG GCTACTAATA CCCAGTCTAC TTTGCGAAhC aATGCTTATC	1020
	CGGCTGTTGA CGAGATGAAT AATTCATTGC AAACCTCCTT TATACTCACT AATGTTTATA	1080
5	TCAATTTTAC ATGACTTTTT AAAAATTAGC TAGAATATCA CAGTGATATC AGCTATAGAT	1140
	TTCAATTTGA ATTAGGAATA AAATAGAAGG GAATATTGTT CTGATTATAA ATGAATCAAC	1200
	ATAGATACAG ACACATAAGT CCTCGTTTTT AAAATGCAAA ATAGCATTAA AATGTGATAC	1260
10	TATTAAGATT CAAAGATGCG AATAAATCAA TTAACAATAG GACyAAATCA ATATTAATTT	1320
	ATATTAAGGT AGCAAACCTT GATATATCAT TGGAGGAAAA CGAAATGACA AAAGAAAATA	1380
15	TTTGATCGT TTTTGGAGGG AAAAGTGCAG AACACGAAGT ATCGATTCTG ACAGCACAAA	1440
	ATGTATTAAA TGCAATAGAT AAAGACAAAT ATCATGTTGA TATCATTTAT ATTACCAATG	1500
	ATGGTGATTG GAGAAAGCAA AATAATATTA CAGCTGAAAT TAAATCTACT GATGAGCTTC	1560
20	ATTTAGAAAA TGGAGAGGCG CTTGAGATTT CACAGCTATT GAAAGAAAGT AGTTCAGGAC	1620
	AACCATACGA TGCAGTATTC CCATTATTAC ATGGTCCTAA TGGTGAAGAT GGCACGATTC	1680
	AAGGGCTTTT TGAAGTTTTG GATGTACCAT ATGTAGGAAA TGGTGTATTG TCAGCTGCAA	1740
25	GTTCTATGGA CAACTTGTA ATGAAACAAT TATTTGAACA TCGAGGGTTA CCACAGTTAC	1800
	CTTATATTAG TTTCTTACGT TCTGAATATG AAAAATATGA ACATAACATT TTAATAATTAG	1860
	TAAATGATAA ATTAAATTAC CCAGTCTTTG TTAACCTGC TAACCTAGGG TCAAGTGTAG	1920
30	GTATCAGTAA ATGTAATAAT GAAGCGGAAC TTAAAGAAGG TATTAAAGAA GCATTCCAAT	1980
	TTGACCGTAA GCTTGTATA GAACAAGGCG TTAACGCACG TGAAATTGAA GTAGCAGTTT	2040
	TAGGAAATGA CTATCTGAA GCGACATGGC CAGGTGAAGT CGTAAAAGAT GTCGCGTTTT	2100
35	ACGATTACAA ATCAAAATAT AAAGATGGTA AGGTTCAATT ACAAATCCA GCTGACTTAG	2160
	ACGAAGATGT TCAATTAACG CTTAGAAATA TGGCATTAGA GGCATTCAAA GCGACAGATT	2220
40	GTTCTGGTTT AGTCCGTGCT GATTTCTTTG TAACAGAAGA CAACCAAATA TATATTAATG	2280
	AAACAAATGC AATGCCTGGA TTTACGGCTT TCAGTATGTA TCCAAAGTTA TGGGAAAATA	2340
	TGGGCTTATC TTATCCAGAA TTGATTACAA AACTTATCGA GCTTGCTAAA GAACGTCACC	2400
45	AGGATAAACA GAAAAATAAA TACAAAATTG ACTAACTGAG GTTGTATTATA TGATTAATGT	2460
	TACATTAAAG CAAATTC AAT CATGGATTCC TTGTGAAATT GAAGATCAAT TTTTAAATCA	2520
	AGAGATAAAT GGAGTCACAA TTGATTCACG AGCAATTTCT AAAAATATGT TATTTATACC	2580
50	ATTTAAAGGT GAAAATGTTG ACGGTCATCG CTTTGTCTCT AAAGCATTAC AAGATGGTGC	2640
	TGGGGCTGCT TTTTATCAAA GAGGGACACC TATAGATGAA AATGTAAGCG GGCCTATTAT	2700

55

	AAACCTAAA GTAATTGCCG TCACAGGGTC TAATGGTAAA ACAACGACTA AAGATATGAT	2820
	TGAAAGTGTA TTGCATACCG AATTTAAAGT TAAGAAAACG CAAGGTAATT ACAATAATGA	2880
5	AATTGGTTTA CCTTTAACTA TTTTGGAATT AGATAATGAT ACTGAAATAT CAATATTGGA	2940
	GATGGGGATG TCAGGTTTCC ATGAAATTGA ATTTCTGTCA AACCTCGCTC AACCAGATAT	3000
	TGCAGTTATA ACTAATATTG GTGAGTCACA TATGCAAGAT TTAGGTTTCGC GCGAGGGGAT	3060
10	TGCTAAAGCT AAATCTGAAA TTACAATAGG TCTAAAAGAT AATGGTACGT TTATATATGA	3120
	TGGCGATGAA CCATTATTGA AACCACATGT TAAAGAAGTT GAAAATGCAA AATGTATTAG	3180
15	TATTGGTGT TCTACTGATA ATGCATTAGT TTGTTCTGTT GATGATAGAG ATACTACAGG	3240
	TATTTTCAATT ACGATTAAATA ATAAAGAACA TTACGATCTG CCAATATTAG GAAAGCATAA	3300
	TATGAAAAAT GCGACGATTG CCATTGCGGT TGGTCATGAA TTAGGTTTGA CATATAACAC	3360
20	AATCTATCAA AATTTAAAAA ATGTCAGCTT AACTGGTATG CGTATGGAAC AACATACATT	3420
	AGAAAATGAT ATTACTGTGA TAAATGATGC CTATAATGCA AGTCCTACAA GTATGAGAGC	3480
	AGCTATTGAT ACACTGAGTA CTTTGACAGG GCGTCGCATT CTAATTTTAG GAGATGTTTT	3540
25	AGAATTAGGT GAAAATAGCA AAGAAATGCA TATCGGTGTA GGTAATTATT TAGAAGAAAA	3600
	GCAATATAGAT GTGTTGTATA CGTTTGGTAA TGAAGCGAAG TATATTTATG ATTCGGGCCA	3660
	GCAACATGTC GAAAAAGCAC AACACTTCAA TTCTAAAGAC GATATGATAG AAGTTTTAAT	3720
30	AAACGATTTA AAAGCGCATG ACCGTGTATT AGTTAAAGGA TCACGTGGTA TGAAATTAGA	3780
	AGAAGTGGTA AATGCTTTAA TTTCATAGAG ATTAGTCGAG GGACCTTTTA CTTATAAAAA	3840
	TGATTTGAAT TAATACTAAA AGATTACAAA GAAGAGGTGG TTTTGTGTGT AAATACAAAA	3900
35	TTGCCTTTTT CTTTTTATGT TAAATCTATA AATTTGAAAC TAAATCAAGG TTAATTCTAT	3960
	GTACACACTT TATATAGGAA GTAGTTTGAA TGTTTATATA ATGTTTACAA AAAAGATGTA	4020
40	GTATTATAAT GTCTAATTC ACATGTGTTT CAGTAAATTT TGTTGTGGAA TGTTAACGAT	4080
	ATACGTATTT TATAAAAaAT TTTTATAAT GATTATTCGA ATGATGCGTA ACGCTTACAT	4140
	CTTATCTAAT GCTAGCTTTT TGACAAAAAT ATGACAATCA ATTAATGTGA TTCTAATAAA	4200
45	TATTCGCAAA TTGCTTTATT GCGATTAAAT TTTTTTGGTG GACTATATA GAAGTTGATG	4260
	AAATATTAAT GAACTTATAT GCAAAAGTAT ATTGAGAAAT AAACAGGTAA AAAGGAGAAT	4320
	TATTTTGCAA AATTTTAAAG AACTAGGGAT TTCGGATAAT ACGGTTCACT CACTTGAATC	4380
50	AATGGGATTT AAAGAGCCGA CACCTATCCA AAAAGACAGT ATCCCTTATG CGTTACAAGG	4440
	AATTGATATC CTTGGGCAAG CTCAAACCGG TACAGGTAAA ACAGGAGCAT TCGGTATTCC	4500

55

	AGAATTGGCA ATGCAGGTAG CTGAACAATT AAGAGAATTT AGCCGTGGAC AAGGTGTCCA	4620
	AGTTGTTACT GTATTCGGTG GTATGCCTAT CGAACGCCAA ATTAAAGCCT TGAAAAAAGG	4680
5	CCCACAAATC GTAGTCGGAA CACCTGGGCG TGTATCGAC CATTTAAATC GTCGCACATT	4740
	AAAAACGGAC GGAATTCATA CTTTGATTTT AGATGAAGCT GATGAAATGA TGAATATGGG	4800
	ATTCATCGAT GATATGAGAT TTATTATGGA TAAATTTCCA GCAGTACAAC GTCAAACAAT	4860
10	GTTGTTCTCA GCTACAATGC CTAAAGCAAT CCAAGCTTTA GTACAACAAT TTATGAAATC	4920
	ACCAAAAATC ATTAAGACAA TGAATAATGA AATGTCTGAT CCACAAATCG AAGAATTCTA	4980
15	TACAAITGTT AAAGAATTAG AGAAATTTGA TACATTTACA AATTTCTAG ATGTTTCATCA	5040
	ACCTGAATTA GCAATCGTAT TCGGACGTAC AAAACGTCGT GTTGATGAAT TAACAAGTGC	5100
	TTTGATTTCT AAAGGATATA AAGCTGAAGG TTTACATGGT GATATTACAC AAGCGAAACg	5160
20	TTtAGAAGTA TTanAGAAAT TTAAAAATGA CCAAATTAAT ATTTTAGTCG CTACTGATGT	5220
	AGCAGCaAGA GGAAGTAGATA TTTCTGGTGT GAGTCATGTT TATAACTTTG ATATACCTCA	5280
	AGATACTGAA AGCTATACAC ACCGTATTGG TCGTACGGGT CGTGCTGGTA AAGAAGGTAT	5340
25	CGCTGTAACG TTTGTTAATC CAATCGAAAT GGATTATATC AGACAAATTG AAGATGCAAA	5400
	CGGTAGAAAA ATGAGTGCaY TcGTCCACCA CATCGTAAAG AAGTACTTCA AGCACGTGAA	5460
	GATGACATCA AAGAAAAAGT TGAAAACTGG ATGTCTAAAG AGTCAGAATC ACGCTTGAAA	5520
30	CGCATTTCTA CAGAGTTGTT AAATGAATAT AACGATGTTG ATTTAGTTGC TGCACTTTTA	5580
	CAAGAGTTAG TAGAAGCAAA CGATGAAGTT GAAGTTCAAT TAACTTTTGA AAAACCATTA	5640
	TCTCGCAAAG GCCGTAACGG TAAACCAAGT GGTTCCTCGTA ACAGAAATAG TAAGCGTGGT	5700
35	AATCCTAAAT TTGACAGTAA GAGTAAACGT TCAAAAGGAT ACTCAAGTAA GAAGAAAAGT	5760
	ACAĀAAAAAT TCGACCGTAA AGAGAAGAGC AGCGGTGGAA GCAGACCTAT GAAAGGTCGC	5820
40	ACATTTGCTG ACCATCAAAA ATAATTTATA GATTAAGAGC TTAAAGATGT AATGTCTTGA	5880
	GCTCTTTTTT GTTTTCAATA ATTGATTCTC TGTAGATATC aaAGTaCTAA CGTTTTAAAG	5940
	GTAAATATT TAATTGGATT GAGATCTGTA TGCGTTATA TCaTTCTGTG TAAATATGGT	6000
45	TCTCCACCAA ATGTGGTGAG TATATAATTT AAAGAACTAT TTTTAAATTA AGAATAATCG	6060
	AACATAAATA AACTTTATGA AATTCAGTA TCATGTTCTT ATAAAAACA ATAGGGCTTT	6120
	TTGctGACGC TAGTGCGGA TAAATAATAA GTTGAATATA AAAAAGATCA CTGCCAATCA	6180
50	TTCGTTTAAAT GGCAGCGATC TTTTTATTT AATTATTTCT CTTTCCACTG CAACATTTGA	6240
	TAACCAATGC GTGGATGTGT TTTAATAATA TCTTTTGCGT CCTCATGACA TTGTGAAAGT	6300

55

CCATATATTC GTTTTAATAT CATCTCATAA GTGAGTACTT TTCCTTTATG ATTTGACAAT 6420  
 AGTTCTAACA AGCTAAATTC ATTTGGCGTC AAATGTACCT CCTGATTATT AATAACAACA 6480  
 5 GATTTGGAGC CAAAGTCGAT GCTTAGCAAA CCGTTAGTAA ATACAATGTT AGTTTCTTGA 6540  
 TGTGACTTAG CGATTCTCTC GATGACTCGT ATTCGTGCCC GAAGCTCATC AACATTAAAA 6600  
 GGTTTAGTCA TATAGTCATT CGCACCGTTA TCTAAAGCTT GAATAATTGT TTGTTCTTCT 6660  
 10 TGTCTTGACAC TTATTACAAT GATAGGAATG TCAGTATGTT GCCTGATTTC TGAAATCAAA 6720  
 CATAATCCAT CTTTATCTGG TAAACCTAAA TCTAATAAAA TGACATCTGG TTTATCAATT 6780  
 TGAATTTTAA AGTGTGCTTG TGTGGCATTG TCGGCTGTAG TTACATTGTA ATAATCTAAA 6840  
 15 GTTAATGCAA CATCAAGTAA ATGTGTGATT GCGTGATCAT CTTCAATTAT CAATATTTTA 6900  
 GATTGCATTA TACGTCTCCT TCGTTAAAGT CTGTATATAT ATTGAAATAG AATATACTGC 6960  
 CGTGTGGTTG GTTCGGTTTA TATTGTAAGT TTGATTGATG TTTGTGTAGG ATAGTCTGTA 7020  
 CTAAATATAA GCCTAGTCCC ATGCTTTTCTT TTTGGTTATC TTTAAAATAT TTATTTGATC 7080  
 CTGTGTAAAA AGGCTCGAAT ATCTTTTGTt GTTCTTCTAA ACTAATTCCA GGTCCCTTCGT 7140  
 25 CTATAACGGC AAATTCGATT TGTTCATAGC TAGCATAACG AATAGATAAA TTGATTTTGG 7200  
 TGTCAGTAGA AGTGTGTTTA ACTGCATTTT CAATCAAATT GAAtAAAgCT TGTAATAATCA 7260  
 ACTTACTGTC AATGTGTATA AACTGTAAAT TTAAGTAGGA TGATACAGTT ATACGCTTTT 7320  
 30 TTAAATGGCG ACGTTCTAAA ATACATATCG ATTTCTTATA CTA 7363

## (2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 10470 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TTAACAATCG ATAACCACAA TACTTCTATT GTAATTGTTT AACGATTTCh CGATTAAAAAT 60  
 45 CATCTAAATC GTCTGGTACT CGACTTGTTA CAATATTGTT GTCTACAcTa CTGACTCATC 120  
 AACTACATGT GCGCCTGCAT TTGATAAATC TTTGCGTACA TTTAATACTG CTGTTAACGT 180  
 ACGACCTTTT AAATCGTCTG TATCTATTAG TATTTGTGGC CCATGACAAA TGGCAAATGT 240  
 50 TGGTACATCA TTTTtagTAA AGTATTTAGC AAATGTGCCA TATCGACCTT CTGTATCTCC 300  
 ACGTAAATGA TCTGGTGAAA ATCCTCCAGG AATTAATAAT GCATCATAAT CTTCTGGTTT 360

55

	ATTTGCAGTA TCTCCAATCA CTACAGTATT AAAGCCTGCA TTCTCTAATG CCTCTTTAGG	480
	GCTTGAATAT TCTATATCTT CAAATTCGTT TGCTAGAATA ATTGCTACTT TTTTAGTCAT	540
5	TGAAAATCAC CTTTCTATAT ATCATTGATA TAATTACTAT AGACAAGTAA ATCAGTGATT	600
	AAACATACAA GATATAAAAA ATATTAAGCG ACTGTCGCGA TATCTAACCC TAACACATCT	660
	TATGTGGCAT TTACTTAGAT ACTAATTTAA CCTTTTCTTC AAGCTGATCT AACAAATCCAA	720
10	TCCATTCAATC TATATCTTCA ACACGTACTT CATCAGGATT TACATGATCG ATATCCTCAA	780
	TAAACTTATT TAAACGCGCT TTTATCTGTT CGATTGTTTG CTGTTCAATC ATAAAAAGTT	840
	AACTCCTTTT ATTTTGTTTT CTTTTTCATT ATTATCCTAA CAGAAATTGC GTTAAAGCGA	900
15	TATAATCTTA GCTATATTTA TGACATTCAA ATTATTTTGA CTTTAAAAAA TCCCCTTTTC	960
	AATTAACATA AATTAAGAGA TAATTTGTGA CGAGTGATAA TACGAaGkGG TaTCATACCG	1020
20	ATATGAACCA AATAGAAAGA AGGAAGTTTA AGACGATGAA TAGCGTCAAA TTGAAGCAAC	1080
	CTGTTAGCAT TTACAATGAT CCATGGGAAG TGAAATTTAT ATACATTTAA ATTTTCATGAG	1140
	ACAATAAACG TTGATTTAAT GCGTTTTTTT GCCTTTTTTA TTTTCCTTAT TTTTTCTGTT	1200
25	TTACAACAAA ATGGTATCAA AAATGGTATC ATTTGTAGTT ATTTTAGCTT CACATATTAA	1260
	AACAACCACA CTCCTAAATT AATAGGTGGT GTGGTTTGTG TGGTTGTGTG GGGATAAAAA	1320
	TAACCGCATC AGTTAAGATG CGGTTATCTA GCAAGGGCCA CGTATTTATA AATACGTTTA	1380
30	GAATCTCTTC GGCAACTTTG CTATAGACAG TCTATGCTGT TACTAAATTA TACCACCACA	1440
	CAAACCTACT CCCATTCAGG AACACAGAGC TTTGTGCTC GTCAGCAACG TCATATGAAT	1500
	TCTCAGTTCA TGTGTGGTG ACACTTTAAA CGGTCTGTGC CAGTAGCGAC CGAGTCATTT	1560
35	CAAGAATGAC CATTTACAT TTATATTATA ACACTTGTCG TCGGTAAGTG TATAGTTTTT	1620
	CAGTTGTATT TAAAGTTAAG TTATCTACTT CGCGCTTTCC TTGCCTTAAT TGTGAAATTA	1680
40	CATATTGCGC TACGCCAGTT TGTTGTGAA TTTGGTAACC TGTATATCA CTTTGATCA	1740
	ATTCAATTAT TTTTAATTTA TAATCACTCA TATTATCTAC GTCCATTCTT TTTATCTAAA	1800
	CAATAAAAAT GTGTCTTTCT CCCGATAAAT AATAACAATG GTAGGCTTAA TAAAAACAAT	1860
45	ATTAAATACA TTTGTTCTGT CATAATTGAA AACCTCCAAA TAATATTATA TTATATAAGT	1920
	GTAAGGAGGA GCCATCAGGC TCCAAGCATA ATGTTAATCT TTGTTGTTTG GCTTTCGGTC	1980
	TAGGTAGCCG AGATGCCaTT CTCTAAGTTG TTTTAACACT TCTGGAATTA TCAGTACTGC	2040
50	CAATACTTGA TGTCTAGAA GTGTTTTTAT TATGTCTAGC ATGAGGCTTT TCACCTCCTT	2100
	ACACATAATT TGTAAGTCAT CAACTAACCT ACAAATATAA TTATACTAAA CAAATGTTTA	2160

55

	GTTATCTACA TTAAATCTT GAGAGAAATG TAAAAAGTT CTAGTAAAT AATAGCACAT	2280
	TTTATCTTTA AATGTAAATA GAAAGCAGGT ATGTAACGCA CCTGCTTAAA TAGaCATGAC	2340
5	TATGTCATT C TAACCTGATT CTCCCCATAA GTCACCTAAT ATCTGATTAG GTGGGGCAGA	2400
	ACCATTCAT GTTCTAATAG GCAAGTAATA ACGTTGCCCC TCCCATGTAT ATCCTACCCA	2460
	AACATGACCA TCTTGTAACA TCACTTCTGT ATAATCACAA TACCCACCAG GTTGGAAGT	2520
10	ATAACCCACT GGACAAGATA AGAATGGCCC CACTTTTCTT ACTGTGATTG GTTGATTGCC	2580
	GTTTGTAAT CTAGCACTTT CTTCCATGTA GTAAGTACCA TATTTATTAC GTTTCATGC	2640
	ACTTGCAACT GGTTTAACTG TATTACTTGA AGCGCTTGAC TCATTAGAGA CAGTGGCAAC	2700
15	CGGTATTTTA CCATCCATGT ACGCCCTAAT CTGCTTGATA AAGTAGTCTT TAAGTTGCAA	2760
	CCGCTGTGCT TCTGGCAATA GACCGCGAGT TACTGGGTCA AAACCAGTGT GTAAAACCGA	2820
20	ACTTCTATGA GGGCATGATG TTGAAGTAAA TTCATTGTGC AATCTGATTG TATTTCTGTT	2880
	TGCTGGTAAT CCCCATTTTT TCAACAATCT AGCGCATTCT TGGAAAGTTG CCTGTTCATT	2940
	TTTTAAGAAT GTCGCGTTAT CTGCGCCCAT TGATTGACAT ACTTCAATAC CGTAATAATA	3000
25	TTTATTACCT ATTTGATTAG CGGTATGCCA ACCTACTTGT GATTCATCTA AGGCTTGCCA	3060
	AACTGTGTTG CCTGATACGT AACTATGCGC AATGCCCCT TCTAATCTTG ATAAAGGTGC	3120
	ATTTACTAAT CCGTTACGAT ATGCTTCAGC AGTCGCCCT TTGCTCCCTG CGTCGTTGTG	3180
30	TATAACTATA CCTTAGGGT TACTACCAG CTTAGGTAGG TCATAACCTT TAACCACATC	3240
	TTTGATGATT TTAAGTTCTA CTGCTTTAGG TTGTGGCTTA GCTGTTTCTT TTTAGGTGC	3300
	TTGTGTAGGA GATTGAACTG ATCGTGGCGC TGTCTCACTT TTAATAATCG GACGGATAAA	3360
35	CCACATAGGG AAATCATAAG CATGTTGTG TCTTGTAAT TTTTCCCAAC CCCAGCCGGG	3420
	TTGTTGCGATT CCGTCAGTCC AGCCACCGCC TAGCCAATTC TGCTCATATA CAATGATGTA	3480
40	ATCTAAAGTT GCTTCAATTA CCCATGCAAC GTGACCATAT CCAGCACCGT AGTTGCTACC	3540
	GAATACCACC ATGTCGCCAG GTTGTGCTAA GAAGTCCGGT GTATTTTGGT ATACAGTAGC	3600
	TAATCCGTCG AAGTTGTTAG CGAACGGAAT ATCTTTTGCA CCTAAACCTT TTAGAAGTAA	3660
45	TCCAAACAAA ACTTTCCAAC CAGCATTGGC ATAATCAAAG CATTGAAATC CATACCATAA	3720
	GTCCACATTG AATTGTTTTT CCTCAGAAGT TTTCAACCAC TCTATAAACT CATTTTTAGT	3780
	TAATTTTGCT TGCAATGTG CCACCTCCAT GATGATACTC ATTCACATCA AAGCCAACAT	3840
50	CGTTAGAGGC GTCTGTGAAA GGTGTGTATG TATCATATTC TTTTGGTGcT TTCGCGCTTA	3900
	ATTCCGGCGT TAAACTACTG TCTGTGTATG ATTTCCACGT AACTTGTTGT TCTTCTTTTT	3960

55



	TTGGGTCAGT AATAACGCCA ATACCTGTAA GTAACGTGAG GATAGCGCCT ATAATTGCGC	4080
	TAGCTTGATT TAATTGAGTA GATAAATCTA ATCCGAATAA ATCCGTGACT TGCTTGATAA	4140
5	ATAGCAACAA TGCTCCAACT AAACCAGTTA GTACTGCTTT GTTTTTGAAT CTCAATTTCC	4200
	AGTTAATATC CATTTGTTTG CTCCTTTTAT CCAAAATAAA AAAACGACTA AAAATTAGTC	4260
	GTTTAAATTT ATTCAATGGT CAATGTCGGA GATCCTGAAT AAACATCACT TATAGTGACG	4320
10	TACAACATCC CTGAAGGATT ACTAAAGTTG ATATTTTTAC TTGCAACTCC GCTATTGACT	4380
	CCTGATATTC CTAAATCACT TGACCCTAAA TTAGTTTGCG AAATCCTCAT TATACCGCTA	4440
15	CGTACATTTT CTATTGTCAC CTGATAACTT TTATTGGGTT CAACTCCATT TATTGTCCAT	4500
	TTTGCTGTTG ATTCTTCTAT GCTATCCGGA TATTTATTTT TAGGTAAGGG TTTTATTACA	4560
	AAAGATGAAG GCTTTTTTCCA TACTTGATA TTTCCAGCAT ATACTTTTGT ATATTCTTCA	4620
20	CCTTCGTAAA TAAACTTCTT TACATTTTAA AAATTACCTT CCATAAAAAT CACCCTTTAA	4680
	TTAAATATAA CGTATTCGGG TCTTTTGTAT ATATATAGTT ATATTCATTT TCTGTTCTCG	4740
	TCCAAATTTT AACCGTCGGT TGAGATGCGC TTTTGTAGTT ATATAAATTA TCCGCTTGTT	4800
25	GTTTAGTAAA AGCTTGAGAT GACAAAACAT ACCGCTCGTC ATGATTATGA TTTTGTGGAG	4860
	CATATAAATC ATTTAGTGTT TGTTTGAATT CCTCAAAATC TTCTGTATTA ACTTTTGAGC	4920
	CAATCTGTTG CAATACACTT TCTGAAATAG AGTTGTTTTG TATTGCTTCT GCTAATTCTC	4980
30	TTAATGTGTT CATAGATTCA GCGCGCTAT CAACTAGTTC AGCAATTTT GTATCCGTAT	5040
	ACGTTTTAGA GTCGTTGAGA GTTGATCTT TGATTTTTTC AACTTCTTGC AATTTATTTT	5100
	CTAACCTTTC AACATTTGCG ATATTGATTT TGTCCAATAA CTCAGGTCTT GCTTTGATAT	5160
35	CTGTATCTTT ACCATCAATT TGCCACATTT TAGTGTGAGG ATTGATTGAT ACTACAGTAC	5220
	CGTTTTTACC GGGTGCGCCT TGTTCTCCTT TTTTACCTGC TTCACCTTTT GCTCCAGGTT	5280
40	GTCCCGGTTT ACCTTTATCA CCTTTCGCAC CTTTAAATCT ACTTTCATTC TTTTCGATGT	5340
	AAGAAATGAC ATCTTTATCT ATTTTCTCTT TAAAGTCTTT GCTCAATAAA TCTGTCGCGT	5400
	TATCTTTTAA AATTCTCGTA ATAGCATCAT CTACCAATTT AACATCGATT TCTTTTGCTA	5460
45	CAGCAGATT C AATACCACTA TCAACGATAT TGAAAGAAAA GTTTGCGACA TGTATTTTFT	5520
	CTTCTTCTTT CTCTAAAAAC AGCTTACAGC GAACATAACC AGCGTGTGTA ATAACCTTTT	5580
	TAGGTATCTT GTAGGTAAGG AAACCTTTTA CAACATCGTC GATAATAAGG GGCTCATTTT	5640
50	TGAATATAGA GCCATCTTCC ATAAACAAAT GTAATCTAGG TGTTAAGCCA TGTGCTTTTA	5700
	GATCGATACG ACCTTGTTTG TCATTGATAC CTATTCTTAT AGATGCTGTA TTTTCATCTT	5760

55

	CAACATCTTT TATTTTGTAC ATTTACACAC CTCTTTATTT ATATTIATCC CTGTGAAGT	5880
	AGATACCTTT TAAGCCGATT TGTATATATA ACTTAGCGAT TGTACTTGCT TGATGTTGGC	5940
5	ACCACTCTAT AGCAGTAGCG TATTGGTGGG TAGCTGGATT CTTAGGATTC CATCTAATTC	6000
	GGTACAATGT GTTTTGACCT TTATTGATGT AATCCTTTCT TACGAAGCTA GCACCGCCCA	6060
	TGATTGCTTT TGCTGGAGAT GTCCAACCTT TATTCCTTGC AAACGTCATT GCGTAGTTAG	6120
10	GATTGTTGTC GTAAGCGCCA ATGCCGAAGT AGTTGTATAC TCCATCTTTT CCGTTAGCGA	6180
	AGTTACTTGT TCCATATCCA CTTTCTAAGA AAGCATGCGC GATTAAATAA ATTCATTAA	6240
	TGTGTGCTT TTTACAAGCT TCTGCGAAGC CTTTACCTTG ATTATTCAAT GTTCCCTTAC	6300
15	CTTTAAGTAT CTTATTAAAGT GCGCTAACTG AAACACCTTG ATACTGCGCT AAATTAAGCA	6360
	TTTGGTAGCA TTGTGTGTTA CTTTCCCAT AAGCTTTTAC ATTCATTGCT GAACTCGTTT	6420
20	GTGCTCGTGT AGCGTTAsec AACCCCAAGC ATTAGATTTT TTCGGGTAC CTCTGCCAT	6480
	TTGTTTATCC AGTGCTTGTT TGAATGTATA AGGACTCGTT TCTGTTATGA TCTGCGTTG	6540
	TTTAGATGCC GAACCATTGT TGGCTGTTGG TGACGAGTCT CTTACATTAG CTATATCAGC	6600
25	GTTTTATTA TCTACCATAA CTTTATTCT AGATTTTGT ACTGTTGGCT TAGTTATAGA	6660
	ATTTAATAAT TTTTCTCTGT TTTTAAATAT ATTAAGTAAT GCCTTTTCTA ATGCTTCGTA	6720
	TTTATCTTTA GGAGGAACAC CGTTGTCAAT CATATTCCAA TTAACATGTT CCAACATTGA	6780
30	ACGCCAAATG CTGTCGTCTA CTTTAAAT TTCAATACTT AGAGGTATCT CATATTTGGC	6840
	CATCATATCT ACAGCTACAA CCATTGCGTG AATCTCATTA AAAATAAAT CATTTTTACT	6900
	CGCACTATAA TCTTCACATA CGTCTATAAC TATATAATCA GGTTCATTAG GAACTTCAAA	6960
35	TACAGCTCTT CTAGGTGCCC AAATATTATG TCTATCAACA TAAAAGTGGG GATATTCTAC	7020
	ATCTGTTTG TATTTCTTCC TACTGTTATA TAACTTTCT ACCGAGCTCA TCGTTGTGC	7080
40	GTTTCTAATC ATTATTCCTT TAGGTTTTTC GAGTCGTCGA TTACCTTCTA CTATAAAGTG	7140
	ATAAATATAT TCTGGATAAT TAACCTCTTG GCTAGAAATA GTGTACTTTA TAGTTGTTAC	7200
	ATCTTTCCAA ATTGGAACCT TTTTATTATT TTTTCGTTA TCATCACTAT CATCTTCTGG	7260
45	TTTAGGTGCC GGTGTAGTTT TGTCTGGATG ATATGGTGGT CTAACAAAAT ATTTAACCCC	7320
	TCCACCTGGT CCATCATGAT AAGAGTGTTT AATTTTATAA GGTGGACTTC CTGTTGCGTT	7380
	ATTTGTATAC CAGTTTTGAT CTACGCCATA CCAATAGTCT TTTGTGCATG GTCCCACTAC	7440
50	AATGTTTACA TGTCTGCCC AACCACCAGT CCAAAACCCC CAGTCGCCTG GTTGTGGTAC	7500
	AAAATCTTTT GTATTTCTAA TTATCTTGAA ATCTCTACCT CTATAATTGG ATTTTGTAGC	7560

55

EP 0 786 519 A2

	TAAATCCCAG	CATTGTGCTC	CCATTCCAGA	ACCAGGTACA	TCAATAGCTA	TTTTGTTTTT	7680
	AGCGATATAT	AACGCCCATT	CAACCACTTC	ACTAGCTGTG	GGCTTTCTAT	TTTTCGGATT	7740
5	AGGTAATCCC	ATGTATGCAC	CTCATTTCOA	TCAAAATAAA	AAGCCAGTGC	CGAAGCACTG	7800
	ACTCTTAACT	GTTATTTACA	TTTACCAAAC	CAGAAGCACG	CCCAGAAGCT	ATATCCTAAA	7860
	ATCCCTTTAA	GCATGGTAAT	CACCTCCTTT	AAATACCAAA	AACAGTTCTT	AGTAAAGCTA	7920
10	TGACAATCGT	ACTGAAGATA	GTCCCTATCA	AACCTAGAAT	CCACATTTTT	ATGTCTCTAA	7980
	TATTCTTGGC	ATTCTTTTCT	TTATTCTTTT	CATCTTCTAC	CTTGTCGCGC	TTTAATTCTT	8040
	CAAAATTTCT	ATCTAATTTG	TCATAAATCT	TTTCTTGCGC	TCTAAGACTA	TCTTCTATTCT	8100
15	TGTCGAATTT	TTCAAACATA	GTCTTATCAT	TTTCTTCTAA	TCGCGTTAAA	CGCCAATCTT	8160
	GTTTCATGTCG	TTTGGTAAAT	CCAAACATTA	TGCCACCCAC	TTTATTCAAA	TTAAAAAGCC	8220
20	ACAAGCATTAA	CACCTGTGAC	TTTTTCATCTT	TGTTTTCTGG	ATATTTTTCT	CCAGTGATTA	8280
	AAGCGTATTCT	TTCTTTIATCG	ATTAAACCCT	TGTCTACGTA	CCACTTAATT	TGCTCGTTTT	8340
	TATAGTAACC	CCAAACATAA	AAAGTTTAA	TGTCTTTAAA	AGTTGGATAA	ATCATCTTCA	8400
25	TTATTTAAAC	GTCCCCCTCA	GTACTTGTTT	TGTTAGTTTT	CAGTTCAGTC	AACTGTTGTG	8460
	TTAACATAGC	GTTTTGTTGA	GCTAATTCCA	TTGTTAATAC	GTTTACTTGT	GCCACCTGCA	8520
	TTTGCATACT	CGCAACCATT	CCGCGAAGTT	CCTCATCACT	TAAATCTGAC	GCACTTTGTT	8580
30	GGTTTGATGC	ATTCCGTACG	TCTTCTTTTT	CGAAATTGCT	ATTGTATTTA	ATTCGCCCGT	8640
	TAGTGAAAAC	AAACTTTCTA	GGTTCGAACT	CTTCTTTAAA	TTTAATAGGC	ACATTGTTAT	8700
	CATCTACATC	TAACTATTG	CGTAAACCGC	CAGTATTAAAC	GAATCCGATA	ACTTCGTTTT	8760
35	TATCGTTTAC	TGTGATTTTC	ATTATTTCOA	CCCCATAATT	TTAGTTATAG	TAACTTTGTT	8820
	GGCAJTCGCT	CCAGAACCTG	ATGTTTTTACC	TAAATCAAAG	TACACATCGT	TATCTATTCT	8880
	TAAAGTAGTG	CTACTTGTTT	TGGATAGTAA	GCACTCATAA	ATACCGCCAC	CGTTGCCGTC	8940
40	TGAGTCAACT	ACATTGCTTT	TACTCAATTG	AATCGCGTTA	GGTAATGCGG	TTAGTCCGAA	9000
	TCCCTCAATA	ACGCCACCTG	GATAAGTTCC	ACTTACCAAC	AAAATAGAAT	AGTTTGTTGTA	9060
45	CGGTTTCAGTT	AGATTGATTG	TTGTACCTAC	ACCATTGTGCG	CCACCGTCGA	ACAATACCGT	9120
	TGATTTATGT	TCATTAGGAA	CTGTCCACTG	TTGCTCAAGT	CTGCCGTTTG	TGATTGATCG	9180
	TGTGTAAATC	TTTTTAGAGT	TATAAGGTGT	GAAGTTAAAT	AGCTTGTTTG	TATCATCTTT	9240
50	AACGAATACC	GATAAATAAC	CCTCATAACT	TTCAACGCTA	CCTGGTAAAT	CCGGCACTCT	9300
	TGTTGCATAG	TAATTACCAG	CAGTTAAATA	TCCCAAATCG	CCTTGCGCAT	TATTTAAGTT	9360

55

GAATTTATCA TCTACATACT GCTTAGCTTG ATTTAAAGCG TTGTTAGACG TTTCTTCAAC 9480  
 AAATTGCTTA GTTAAGTTTC CATCATTCTT TTTATAAAAC GGGTACCATG TGCCGTAGAT 9540  
 5 TTTGTATTTT GTGTACTCAT CGTTTGAATC GTCTGGGTAC CATGTTGCAC GAGCAGTATT 9600  
 ATTATCAACA ACATAAACAA CTAACACACC AGATTTGCTT GATGTATAAG TTGATTCATC 9660  
 GAACGAAGAA CCGTCATCAA CACCATCTTG TCCAGGCTTC TCTAACGTGC CTATATCCGT 9720  
 10 CTTTCTGGC GCATCTGTTG CATTAGTAAT ATGAATAATC CTAGATGTGT TAACTGCGCT 9780  
 TAAACGCTA TCTATGGACT GCTCATACTA TTCAATTGCT TTACCGTAAT CATCTGTAAG 9840  
 TTTAGACTTT TGCCAATTCG TTGTTGAATT ACCTTTAACA AGGTCAGCGC CATTGATTTG 9900  
 15 TTGTTCAACT TCGTTAACAC GTTCAAAAAT CGCTTGCTCT TTTTCAACTA TTTTATCGAA 9960  
 TTCAGCTGTA ACAGCTTGTG TTGCACTAGT TTGCGTCGCA GTAATAGCTT GTATAGCTTC 10020  
 GTTTTGCTTG ATTTGCTTTT GTTGAATGCC TTTTGTGCGA CTATCATTC A TTTTGCTAT 10080  
 20 TAACGTTTGT GTATCAGCCA TATTTTGCTT TAATTGGTTA AAATCTTTAC CGACAGCTTC 10140  
 GATAGTATCT TGAATAGATT TGATATAAAC AAGCTTTGTT ATACCATCAA ACCCACTAAC 10200  
 25 TAAATCATTT TCAATATTGA AGCTAAATTG ACGTTCAACA ACAACATTAT TACTCCCGTT 10260  
 TTGTGTAAAG AATGCCTGAG CATGCACCTT GCCTGAATGT TTTAAAAATT CATTCCGGTAT 10320  
 CACATACTGC AAACGCCCAT TAATTGCGTC TACTATCGTT AATTCGCTG AAATATAAGC 10380  
 30 GCCTCTATCT ACGTTATAAT CATCGGTTTT TAAACGATA GATGTTTTAA CATGTTTACA 10440  
 ACTTATAGAT AAGGGTCTGT TATnCTTAGT 10470

## (2) INFORMATION FOR SEQ ID NO: 21:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3647 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

45 ATCAGATCTT GAGAATCGAG TTATTAAGTC TATCGAAGAC TTAATAAAA TCCAACCATT 60  
 CATGCCTACA CAAGATTTTG ATTTTAAAC TAAAGAAATT CAATCAAACA TTTCTGAAGA 120  
 AAGATTTATC GAAATGATTC AGTATTTCAA AGAGAAAATA ACAGAAGGGG ATATGTTCCA 180  
 50 AGTTGTGCCA TCAAGAATTT ACAAATATGC GcATCATGCT AGTCAGCATT TAAATCAACT 240  
 TTCGTTTCAA CTGTATCAAA ATTTAAACG ACAAACCCA AGTCCATATA TGTATTATCT 300

55

## EP 0 786 519 A2

	TCAAATTGTA	ACAACTAATC	CTATTGCAGG	TACGATTCAA	CGTGGTGAGA	CGACACAAAT	420
	AGATAATGAG	AATATGAAAC	AACTACTTAA	TGATCCAAAA	GAATGCAGCG	AACATCGTAT	480
5	GCTAGTTGAT	TTAGGACGTA	ATGATATTCA	TAGAGTAAGT	AAAATCGGTA	CCTCAAAAAT	540
	TACTAAATTA	ATGGTTATTG	AAAAATATGA	ACATGTTATG	CATATCGTAA	GTGAAGTCAC	600
	AGGTAAAATA	AATCAAAATT	TATCGCCAAT	GACAGTTATT	GCGAATTTAT	TACCAACAGG	660
10	TACCGTTTCA	GGTGACCAA	AATTACGTGC	AATTGAAAGA	ATATATGAAC	AATATCCACA	720
	TAAACGGGGC	GTTTATAGTG	GTGGTGTGG	ATACATAAAT	TGTAATCATA	ACTTAGATTT	780
15	TGCATTAGCA	ATTGGAACGA	TGATGATAGA	TGAGCAGTAT	ATCAACGTAG	AAGCTGGTTG	840
	TGGCGTTGTA	TATGATTCTA	TTCTGAAAA	AGAACTGAAT	GAAACGAAAT	TGAAAGCTAA	900
	AAGCTTATTG	GAGGTGAGCC	CATGATCTTA	GTTGTAGATA	ATTATGATTC	CTTTACATAT	960
20	AACCTAGTGG	ATATTGTTGC	TCAACATACT	GACGTCATTG	TTCAATACCC	TGATGATGAT	1020
	AATGTGCTGA	ATCAATCGGT	GGACGCTGTT	ATTATATCTC	CTGGTCCAGG	GCATCCATTA	1080
	GACGATCAAC	AGTTAATGAA	AATCATATCA	ACCTATCAAC	ACAAACCCAT	TTTAGGTATT	1140
25	TGTTTAGGGG	CTCAGGCACT	GACTTGTTAC	TACGGTGGAG	AAGTCATTAA	AGGCGACAAG	1200
	GTTATGCACG	GCAAAGTTGA	TACACTAAAG	GTTATATCGC	ATCATCAACA	TCTGTTATAT	1260
	CAAGATATAC	CAGAACAGTT	TTCAATTATG	AGATATCATT	CATTAATAAG	TAACCCTGAC	1320
30	AATTTTCCAG	AAGAATTGAA	AATTACTGGA	CGTACCAAAG	ATTGTATACA	GTCATTCGAG	1380
	CATAAAGAAA	GACCGCATTG	TGGTATTCAG	TACCATCCTG	AATCATTTGC	TACAGACTAT	1440
	GGTGTCAAAA	TAATTACAAA	TTTCATTAAAT	CTAGTGAAGG	AAGGATGAAA	ACCATGACAT	1500
35	TACTAACCAAG	AATAAAAACT	GAAACTATAT	TACTTGAAAG	CGACATTAAA	GAGCTAATCG	1560
	ATATACTTAT	TTCTCCTAGT	ATTGGAAC TG	ATATTAAATA	TGAATTACTT	AGTTCCTATT	1620
	CGGAGCGAGA	AATCCAACAA	CAAGAATTAA	CATATATTGT	ACGTAGCTTA	ATTAATACAA	1680
40	TGTATCCACA	TCAACCATGT	TATGAAGGGG	CTATGTGTGT	GTGCGGCACA	GGTGGTGACA	1740
	AGTCAAATAG	TTTCAACATT	TCAACGACTG	TTGCTTTTGT	TGTAGCAAGT	GCTGGcGTAA	1800
45	AAGTTATAAA	ACATGGtAAT	AAAAGTATTA	CCTCaAATTC	aGGTAGTACG	GATTTGtTAA	1860
	ATCAAATGAA	CATACAAaCA	ACAACTGTTG	ATGATACACC	TAACCAATTA	AATGAnAAAAG	1920
	ACCTTGtATT	CATTGGTGCA	aCTGAATCAT	ATCCAATCAT	GAAGTATATG	CAACCAGTTA	1980
50	GAAAAATGAT	TGGAAAGCCT	ACAATATTAA	ACCTTGtGGG	TCCATTAAAT	AATCCATATC	2040
	ACTTAACGTA	TCAAATGGTA	GGCGTCTTTG	ATCCTACAAA	GTAAAGTTA	GTGCTAAAA	2100

55

	AAGCAACACT ATCTGGTGAT AATTTGATAT ATGAATTGAC TGAAGATGGA GAAATCAAAA	2220
	ATTACACATT AAATGCGACT GATTATGGTT TGAAACATGC GCCGAATAGT GATTTTAAAG	2280
5	GCGGTTCAACC TGAAGAAAAT TTAGCAATCT CCCTTAATAT CTTGAATGGT AAAGATCAGT	2340
	CAAGTCGACG TGATGTTGTC TTAATAATG CGGGTTTAAG CCTTTATGTT GCAGAGAAA	2400
	TGGATACCAT CGCAGAAGGC ATAGAACTTG CAACTACATT GATTGATAAT GGTGAAGCAT	2460
10	TGGAAAAATA CCATCAAATG AGAGGTGAAT AATATGACGA TTTTATCAGA AATTGTTAAA	2520
	TATAACAGT CACTTTTACA AAATGGCTAT TATCAAGACA AACTTAATAC CTTGAAAAGT	2580
	GTGAAGATTC AGAATAAAAA ATCTTTTATA AACGCAATTG AGAAAGAACC AAAGCTAGCA	2640
15	ATTATTGCAG AAATTAAATC GAAGAGTCCT ACAGTTAATG ACTTACCTGA ACGAGATTTA	2700
	TCGCAACAAA TCTCAGATTA TGACCAATAT GGTGCAAATG CCGTGTCCAT TTAACTGAT	2760
	GAAAAGTACT TTGGTGGTAG TTTTGAAAGA TTACAAGCAT TGACGACAAA AACACATTA	2820
20	CCCGTATTAT GCAAAGACTT TATTATAGAC CCGCTTCAA TTGATGTTGC TAAACAAGCT	2880
	GGTGCATCTA TGATTTTATT GATCGTTAAC ATCTTATCTG ATAAACAATT GAAAGATTTA	2940
25	TATAACTACG CTATATCGCA AAATCTAGAA GTGTTAGTTG AAGTACATGA TCGCCATGAA	3000
	TTAGAACGTG CCTATAAGGT TAATGCTAAA TTGATTGGTG TAAATAACAG GGACTTAAAA	3060
	CGATTTGTTA CAAATGTGGA ACATACAAAT ACTATTTTAG AAAATAAAAA AACAAATCAT	3120
30	TATTATATTT CTGAAAGTGG TATTCACGAT GCATCTGATG TAAGAAAAAT CTTGCATAGT	3180
	GGTATCGATG GCTTACTAAT AGGTGAGGCG CTTATGCGTT GTGACAATCT ATCTGAATTT	3240
	TTACCACAAC TGAAAATGCA AAAGGTGAAG TCATGATGAA ATTGAAATTT TGTGGCTTTA	3300
35	CATCAATAAA GGATGTTACA GCGGCCAGTC AATTACCTAT TGATGCGATA GGTTCATCC	3360
	ATTATGAAAA AAGTAAAAGG CATCAAACAA TTACCCAAAT AAAAAAGTTA GCGTCTGCTG	3420
	TTCCAAATCA TATCGATAAA GTATGTGTCA TGGTAAATCC TGATTTAACA ACAATTGAAC	3480
40	ACGTATTAAG CAATACGTCA ATTAACACAA TACAGTTACA CgGCACAGAA TCTATTGATT	3540
	TTATACAGGA AATTAAAAAG AAATATTCAA GCATTAAAAT CACTAAAGCT TTAGCTGCaG	3600
45	ATGgAAAACm TwATCCCAA caTtAAtnAA tnTTAgGGGG TCCGTGG	3647

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5966 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

	CcAcCTTGAC CACCTTTACG TGGAATCTTT TcmCCTkGAG CAACaTCGaT AATaTATATT	60
5	GAAAgtCAAC AAGTTCTGGA CTAAATGTTG CTGCTAAGTT ATCGCCACCA GATTCTATGA	120
	AAATTAGTTC TATATCGTCA TGACGTTCTA ATAATTCGTC TATTGCTGCA AAGTTCATAG	180
	ATGCATCTTC ACGAATCGCA GTATGAGGAC ATCCACCAGT TTCAACACCA ATGATACGAC	240
10	TTTCAGGTAG AACTCCTGAA TTTACTAATA TCTTTTCGTC TTCTTTTGTA TATATATCAT	300
	TTGTAATAAC GCCGATACTC ATTTCTTTTG AAAGACGTTT TACAACTTTT TCAATTAATT	360
15	GTGTTTTACC TGCACCTACA GGACCACCAA TACCAATTTT AATCGGATTT GCCACAATTA	420
	TAACCTCCTA TGATATGAAA tTCTAACATT GaCGTTCTCA TCGGCCATTT GATTTAGTTC	480
	TAAACCAGGC GCTGTCATGC CAAAATCTGC TTCTTTTAAT TCGAAAATCT GCTTTCTTGT	540
20	TCCTTCTATA TAAGGAATCA TGTGAGTAAC TATCTTTTGA CCAGCAGTTT GTCCAAGTGG	600
	AATAGCACGA ACAGCATTTT GAGTTAAACT TGAACATTTT TGATATAAAT AGTAATCAAT	660
	AATCGTTTTCA ATATCTACAC CTAAATGATG GCCTAGCATA GTAAAACAAA TAGCTGGATT	720
25	TnACTTTGCT TTCTTATCTT GCATTTGTTG ATGATACCAA GCAATCCATG GGCTATtATA	780
	AAGTTCTAAA GCCAATTTAA CCATGCGAGT CCCCATTTGT kTTGCACCAA CACGTGTTTC	840
30	TTTAGGTAAG TTTTGrACAr ACATCAGTTT ATCTATGTGT AATACTTTTT GTGTATCATC	900
	ATTTTCCAAT GCATCATAAA CTaaACGCAT GGCTAAACCA TCAGAATAGG TAAGTTGCTC	960
	TTGTAAAAAC ATTTTTAACC AAGCAATAAA AGTATGATCG TCATGAATTA TATTTCGTTG	1020
35	AATATATGTT TCAAGACCAA ATGAATGACT GAAAGCACCT GTTGAAACT GTGAATCACA	1080
	GAACTGAAAT AATCTTAAGT GTGTATGATC AATCATGAGA ATGCCCTATA TGTCTGAAAG	1140
	CCTTATTAAC TTTACGGTCT TCTCGAACAT ATGGGATGCC TAAACTTTTT AATAAATCTT	1200
40	CAACTAAATA ATCATATTGT ACTAGCATTT CAGTCTCTGT AAATTGTGCT GGCAAATGAC	1260
	GATTTCTTAA TTGATGGGCT ATATCTCCCA TTTCTTGCAA TGTCTTGGT TGAATCACTA	1320
	AAAGATCTTC TGAATTAACA TCCACAATAA TCATATTATG GTCATCTGCG TATAAAATAT	1380
45	CTCCATATTG TAAGTCAATA GGTTGTTTTA AACGAATGCC TATTTCAAGT CCATGGTCTG	1440
	TAACGACTCT TTGAATACGT TTAACAAGAT CTGAATTTTC AAGGTATACT TTTTCGACGT	1500
50	GCTTTTGTTT TTCTGAATTT GACAAATGG CAATATTGCC TTGGATTTCT TCAACAATCA	1560
	TTCTATGTTT CTCCTAGAAT AAGAAGTATC TTTGAGTTAA TGGTAACTCA GTTGCTGCAT	1620
	TACTTGTAAT TTTTCTCCA TCTACATATA CTTCATATGT TTGTGGATCA ACGTCTAATT	1680
55		

	GACGCACCAT GCGTTTTAAA TTTAATGCAC GATTGATACC ATTTTCATAA GCAGTTTTAG	1800
	ACACGAATGT CATTGACGTA CTTGTAAGGT TTCCGCCGTA TTGACCATAC ATTTTACGGT	1860
5	ACTTCATCGG TTCAGATGTA GGTATAGAAC CATTTGCATC GCCATTTACG GCAGAGTTAA	1920
	TTAATCCGCC CTTTACAAC T AATTCAGGTT TAACCCCAA GAAAATTGGG TCCCATAAGA	1980
10	CAATGTCAGC TAGTTTGCCC GGCTCGATAG ATCCTACATA TTCAGAAATA CCATGTGTAA	2040
	TTGCTGGGTT AATTGTATAT TTAGCGATAT AACGTTTGAT GCGATTATTA TCATTATGTT	2100
	CAAAATCACC ATCTAAAGGA CCACGTTGTT CTTTCATGCG ATGTGCTACT TGCCATGTTT	2160
15	GTGTAATTAC TTCACCTACA CGGCCATTG CTTGTGAATC GGAACATACT ATACTGAATA	2220
	CACCCATATC TTGCAGACA TCTTCTGCTG CAATCGTTTC TTTACGAATA CGTGAATCTG	2280
	CGAATGCGAT ATCTTCAGGA ATAGCCGCAT TTAAATGGTG AGTAATCATT ACCATATCTA	2340
20	AATGTTTCATC TACAGTATTA TGTGTATAAG GCAAAGTTGG ATTTGTAGAT GAAGGTAAAA	2400
	TATTTGAAAA TGCAGCGGAT TTAATTAAAT CAGGCGCATG ACCGCCACCA GCACCTTCAG	2460
	TATGGTACAT ATGAAGTACA CGGTCTTTAA CAGCAGCCAT TGTGTCTTCC ATAAATCCTG	2520
25	CTTCATTTAA AGTATCTGCA TGTAATGCAA TTTGAACATC AAATTCATCA GCAACATCTA	2580
	ATGCATGACT CAAAGCAGAT GGTGTTGCAC CCCAGTCTTC ATGTACTTTT AATCCAATTG	2640
30	CTCCGGCATT GATTTGTTCA ATGAGTGCAG TTGGATTTGT TGCTTGTCTT TTACCTGTAA	2700
	AACCGACATT AATCGGTAAA CcTTCCGGCAG CTTCTAACAT TCTATGAATA TGCCATGGAC	2760
	CTGGAGTTAC AGTTGTTGCT TTAGAACCCT CTGAAGCACC AGTACCACCA CCAATATGAG	2820
35	TCGTAATACC ACTTTCTAAT GCGACCTCTG CTTGTTCAGG ATTAATAAAA TGAACATGAG	2880
	TATCAATACC ACCAGCAGTG ACGATTTTAC CTTCAGCGGC AATGATATCT GTTGTGTAAC	2940
	CTATAATAAT GTCGACATTA TCCATTATAT CTGGGTTGCC GGCATTACCT ATGGCGAAAA	3000
40	TATAACCATT TTTAATGCCT ATATCAGCTT TAACCACTTT ATCGTAATCG ATAATAACGG	3060
	CATTAGAAAT GACAAGGTCT GCAACGTTCA CGTCATCACG TGTTACACGA GGATTTTGCG	3120
45	CCATACCGTC TCTAATAGAT TTACCACCAC CAAAAGTAGC TTCTTCACCA TAAACCGCAT	3180
	AGTCTTTTTT TATTTGAGCA AATAGATTCG TATCACCTAA ACGAATGGAA TCTCCAACAG	3240
	TTGGACCGTA TAAGCTCGTA TATTGATTTT GCGTCATTTT AAAGCTCATG ATCTTTTTCC	3300
50	TCCTTTTTTA TTCACGTTTT CAGCACCGTT ATCTCCGAAT ACACCTGCAT ATTCATCATT	3360
	TTCATCAGTT GGGCGATAGA CACGTGACTC ATCGATAGGA CCATTGACCA TACCACGAAA	3420
55	ACCAAAAATT TTACGTTTGC CAGCATATTC AACTAATTGA ACTTCTTTTT TATCCCCAGG	3480



	TTCGAAATCT	AATGCTGCAT	TTGCTTCATA	AAAATGAAAA	TGTGAGCCCA	CTTGAATTGG	3600
5	TCGATCTCCT	GTATTTTCAA	CTTCGATAAC	TGTTTCAGGA	TGATGGTTAT	TAATTTCAAC	3660
	CTCTGTACTT	TTTGTAATAA	TTTCTCCTGG	TATCATTTGA	CTGCCTCCTT	TAAACAATAG	3720
	GGTGATGTAC	TGTGATTAAAC	TTAGTACCAT	CGGGGAACGT	AGCCTCGATT	TCGATATCTG	3780
10	TAATCATGTG	TTCGACACCA	TCCATGACAT	CTTCTTTGTT	TAGAATTTGT	CTACCATAAC	3840
	TCATTAACTC	TGCAACGGTC	TTACCATCGC	GTGCACCTTC	TAATAATTCA	TCGCTGATTA	3900
	AAGCTAATGC	CTCAGGATGA	TTTAGTTTCA	AACCACGTGC	TTTACGACGA	CGTGCAACTT	3960
15	CCGCCGCCAC	TACAATCATT	AATTTGTCTT	GCTCTCGTTG	TGTAAAATGC	AAATTAAAAC	4020
	CCCCAATTTT	ATATTAGATA	CaATTTACAA	AATTTATATT	AATCCTAATT	GTTGTGATAA	4080
	ACAAGTAATA	TACAAAGTTC	AATGTGTAAT	TAGAAAATTA	TATTTTTTAGC	ATATCCGATA	4140
20	TTGAAGCAAA	CAATCTAATC	GAAAACAAAT	AGTGGAATAT	ATTTATGTAA	AAACCAAAAT	4200
	AGTTTTTAAT	ATAACTTTTC	ATAGAATAGT	AGTATATTAA	TGAGTAATGA	TTCAAAGGAA	4260
25	AGGTGAAAGA	TTTGAAGATA	ATAGATGTGC	TTTTGAAAAA	TATATCTCAG	GTTGTGTTAA	4320
	TTAGTAATAA	ATGGACAGGA	TTATTTATCT	TAATAGGATT	ATTTGTAGCC	GATTGGACAA	4380
	TTGGATTAGC	GGCTATTGTA	GGTAGCATCA	TCGCCTATAC	TTTTGCGCGT	TTTATAAATT	4440
30	ATAGTGAGGC	AGAGATTAAT	GATGGGTTAG	CTGGATTTAA	TCCAGTGCTA	ACTGCCATTG	4500
	CGTTAACAAT	CTTTTTAGAT	AAGTCAGGAT	TAGATATTGT	TATAACAATG	ATAGCAACTT	4560
	ATTAACGTT	ACCAGTTGCT	GCTGCAGTGA	GAGAAGTTTT	AAGACCATAT	AAAGTTCCGA	4620
35	AGCTGACGAT	GCCTTTTGTC	ATTGTGACTT	GGTTTACAAT	TTTACTTTCA	GGACAGGTTA	4680
	AATTTGTAGA	TACATCGTTA	AAGTTAATGC	CTCAAAACAT	TGAAACGGTT	AATTTTAGCA	4740
40	CAATGATAG	AATaCATTTC	ATTCAGTCAT	TATTTGAAGG	ATTCAGTCAA	GTATTTATCG	4800
	AAGCGAGTGT	AATTGGTGGC	GTATGTATTT	TAATCGGCAT	ATTGATAGCA	TCAAGAAAAG	4860
	CAACACTCTT	AGCTGTTATA	GCTAGTTTGT	TAAGCTTTAT	CATTGTAGCT	CTATTAGGTG	4920
45	GTAATTATGA	TGATATTAAT	CAGGGATTAT	TCGGTTATAA	CTTTGTATTA	ATGGCAATCG	4980
	CACTAGGATA	TACATTTAAA	ACAGCGATTA	ACCCTTATAT	TTCGACTTTT	TTAGGTGTGT	5040
	TATTAACAGT	AGTGGTGCAA	CTAGGTACAA	CAACATTGCT	TGAACCGTTT	GGCTTACCTG	5100
50	CATTAACATT	GCCATTTATT	ATCGTGACAT	GGATTTTATT	ATTTGCTGGT	ATTAAACATG	5160
	ACAAAGTAGA	TGCTTGATAG	TTAAATCAAA	CCTAATATTG	TTTGAATATC	ACCTTAAACT	5220
55	ATACAGCGAA	TTGTATAGTT	TAAGGTGTAT	TTTTATGGAT	AAAATTAAGT	GCATACTTAA	5280

GTGTTAACT AGGAATAAAT AATTTATATT GTGTGTTGTG TGGGGTGAAT AATATGAATG 5400  
 ATATGGATAA TTCCTTTTAA ATAACAACGG AAATTCAAAG AAAATGGATT GAAAAATTCA 5460  
 5 AAGTAATTAG AGATACATTT AAGGCTAAAAG CTGAATATAA TGATCAACAT AGCCAATTTT 5520  
 CATATAAAAA TATTGAATGG TTAATTAAAG AAGGTTATGG AAAATTAACG TTACCAAAAG 5580  
 CATATGGTGG TGAAGGTGCG ACCATAGAAG ACATGGTTAT TTTGCAATCA TTTTGTAGCG 5640  
 10 AACTTGATGG TGCCACAGCA TTATCTATTG GTTGGCATGT GAGTGTCGTA GGACAAATTT 5700  
 ATGAACAGAA ATTATGGTCT CAAGATATGT TGGAGCAATT TGCTGTTGAA ATTAATAATG 5760  
 15 GTGCATTAGT TAATAGAGCA GTTAGTGAAG CTGAAATGGG TAGTCCAACA AGAGGGGGAA 5820  
 GACCAAGTAC ACATGCTGTT AAAGCTGATG ATGGGTATAT TTAAATGGT GTGAAGACAT 5880  
 ATACATCAAT GAGTAAAGCA CTAACACATA TTATTGTTGC TGCTTATATA GAAGAATTAG 5940  
 20 AAAGTGTGG TTTTTCTTA GTAGAC 5966

## (2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 17310 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CTGTGTCATC GCGAAATAGT TAGGGTCATT CATTAACTCT TTTGAACGTA TTTCATCAAA 60  
 35 ATATAACAAT TTCATTAGTA AAGGGGACTT GTTCAAACCA GCTATAATAC AAAATAGACC 120  
 TATAGTCACA CTGCTTATAA TATAAGAGGT AACGATCACT TTTTGTCTAT TACCTAACTT 180  
 AAAGGTGATC ATCCCTAAAT AGAAATAAAT GACTACAAAT GCATATTTAA CTGTAGATGC 240  
 40 AAGAACTTCC TTAACCGTAA TAAATATCAA ATCATCAAAA AATaGCaAAC AArGCGTAAT 300  
 AATCATACGA TATGTATACA AAATAATGA<sub>m</sub> AACTGT<sub>m</sub>AA AAATGATTTG CCTTTAATAA 360  
 ATGGTTAGCG AAAAACAGTA AATAAACTAA TATTAGTAAT GTGATAAAGT CAGCTATAGA 420  
 45 AACATTCACA CCGGCAATAA CCGAAGATTG CTGAATAAAA ACCGCTAAAC CGATAAGTAA 480  
 CAATGTTAGT AATTTACTAT TGTGTTGATT TTCCATTATA AACGTCTTCC ACTTCTTTAA 540  
 50 TCATTTTCTC CTCAGTAAAA CATTCTAAAT AACGTTTTCT AGATTGATTA CTCATTTTGA 600  
 TGTAATCACT GTCTATTAAA TATTTTCCA GGACTTTAGC AATAGTTTCG GGTTGGTTGT 660  
 TCATCATACA TATACCATTA TCAGCTACTA ATTCTGAAAT ACCGCCAACA TGAAGGCTA 720

55

	TTATTAAAAT AAACGTATCG TATTGTGATA ATAAATGACT CGCATTAAATG ACATTGCCCA	840
	AAAATGTGAC ATCATTTTCT AACCCAGCTT GTACAACCTG TTGCTGACAA TCATTTAATG	900
5	TAGGTCCATC GCCTATAAAT GTAAAATGCG CATGATTACT GTTATGTAAT TTCAATATCT	960
	CTATTGCCGC GATTAGATT TGTGGCAATT TTGGATAAGC AAATCTTGCA ATCATAACAA	1020
10	ATTGATGCTT TGTGCGGGCA TTAATCTGTA AATCTTGTTT ATTAGGCAAC ATTCCAACCTA	1080
	CTTCGCCAAT ATTGTTATGT GATTGGCTTT TTAGCGTTTG CTTAACAGCG GGAACATCTG	1140
	CAATACCATT ATGTATTGTG GTTAATTTCA ATCGATTAAA TCGATATTTT AACGCTAACT	1200
15	GTTTATCGAA ATCTGAAACA CAAATAATGC TATCTGTAAT AAGTGACATT AATTTTTCGA	1260
	TAACTAAATA TAGAAATTTT TTAGCTGGTT TAACACCCTC TGTAAAAGCC CATCCATGTG	1320
	CAGTAAAAAC TATACGTGTG TCTTTCGATT TCGAAATGAa CTtCGCAATT CGTCCGACCG	1380
20	TtCCAGCTTT GGAAGAATGT AAATGGATAA CATCAGGTTT AATTTTCGAG AATAACTGTG	1440
	CTAACACTTT GACAGCTAAA ATATCTTGTT TAAAGTCAAT TGGACCTACT AAATGTTCTGA	1500
25	TAATAATTAC ATTAACCTCT GCATCTAGTT GTTCAATCAT TGGTCCATGA TTGCCTACAA	1560
	TGACATAAAC ATCATTGTGT ACGCAAAAAT GGTGGCGAG TTGAATGAGA TGTGTTTGTG	1620
	CACCACCATT GTCTGCTTTA GTAATACAAT ATATAATTTT CAACTGTTAC AAACCCCTTT	1680
30	AATGCTATAC TTTCAATTTT TTAACATGGC TATCTCATCA GATGAATAGT ATTTATAGCC	1740
	ATGCAAATCA ATGATGGCAC ATATTTCTTA ATGCCATTTG ATACTGTCTC AAGGGATTCC	1800
	TCGTTATACT GTAACAATTG GTCACAATCT TTAAAATATA ACTTTTATTT GAACTTATTA	1860
35	AGTAAATTAA GACTACCTTG AGCCTTCCCC TGTAATAACA ACCATCAATG TTCTAATTGA	1920
	TATATATAGT TCCATCATT AACTACCTTT ATGTATATAT TTCATGTCAT ATTTAGTTT	1980
	TTGTTCGGT GTTAAGTCAT ATCCACCTTG AATTTGCGCA AGTCCTGTTA ACCCTGGTGT	2040
40	AACAAGACAT CTTTGCTCGA AACCTATCAC TTCTGAACTA AATAATTCTA CAAATTCCGG	2100
	ACGTTCCGGG CGTGGTCCAA TAAAACTCAT TTCCCCTTTA ACAACATTAA TTAGTTGTGG	2160
45	TAATTCATCA ATGCGTGTTT TACGAATAAA CTTCGCGACA TTGTTATAC GATCATCATC	2220
	TTTATCAGCC CATTGCGCAC CGTTTTCTC TCGGTTTTTG CACATCGAAC GTAATTTGTA	2280
	TATTTTAATT AATTTACCCA TCTTCCAAC TCTAACCTGA CTATAAATAG GGTTTCCTGG	2340
50	CGAATCTATG ACGATAGCAA TGGCGAATAT AACCATAATC GGTAAAGTTA AAAATAATAA	2400
	AACAATGCTT AAAATTAAGT CAATCGCACG TTTAATTGGG TAATAGCTTT TTCTCACTTC	2460
55	TTCTAGTTTG TCTAATTTTC TTTGATAGGC ATAACCCTTA TTATTATGGA CAGCTTCAAT	2520

	AATTAAAGTA ATCCTTTAAA CCTGTTTCTA CTGTATATTT AGGAACAAAT CCTAATGCCT	2640
	TTAAGTTAGA AATATCTGCA TAAGAATGCT TAATATCTCC TTTTCGTGCT TCTTTAAATT	2700
5	CATGCTCGAC TGATTTTCCA TATAATTCAC CAATAATACG ATAAACCTCT AATAAATTAG	2760
	TAAAAGTGCC TGTACCAATG TTATAACCGT GTCCAATTGC ATCTTTGTGT TCCATAATTA	2820
10	AGCGTACAGA TTGAACAACA TCATATACAT ATACAAAATC TCTAGTTTGC AGTCCGTCAC	2880
	CAAAAAATGT AAATGGCTTG TTATGCTCAA ATGAATCGAA CATCTTTGAA ATCACACCTG	2940
	AATATTGTGA CTTAGGATCC TGTCTTGGCC CAAATACATT AAAAAATTTA ACAACCGCTG	3000
15	TTGGTATGTT ATATAACGAA CAATAATTTA ATGTCGTCCG TTCGCCGTAA TATTTATCTA	3060
	TTGCATATGG TGATAATGGT AAGATTAATG ATTGATCACT TTTAGGCAAA TCAGGAAGAT	3120
	CACCATAAAC AGCTGCTGAC GAAGCAAAGA TAAAACGTTT TATATGATTA TTATATTTTT	3180
20	TAATGATTTT TAACAATCTT AATGTTGCTA CGACGTTTAT TTCTTGAGAT AAGATAGGTT	3240
	TCTCAACCGA CTCAGCAACA CTAACCTAATG CTGCTAAATG AATAACATAA TCAAATTGAT	3300
25	ATGTCTTCAT GATTTGTTCA ACTGCATCAT ATTCACGAAT ATCTAATTCA AACACATGAT	3360
	CGTCAGCCAA ACTTTTAATA TTTTCTCGTT TACCTGTTCT ATAGTTATCT AGAACATAAA	3420
	CATCATAATC TTGTTGTAAA TCATCTACTA AATGCGACCC AATAAAACCA GCCCCACCAG	3480
30	TTATCAAAC TCTTTCCAAA TCTTCCACCT CATTTATACA TTAAAAATAT ATCATAAAAA	3540
	CATAAAGTAT TGTAAGCTTT TTATCGATAT TTTTATTTA TAAAAATAAA ATGAGATAAC	3600
	TTTGTGAATT TTTATTGAGA TAAATTAGAT AGTGGTGTTT TTGTGATGTT TTATAATATC	3660
35	TTGGGTGTGT TAATACTAAT AATGCTTTCA ACTGATGCAT TAGACTGTGA CATCATAACT	3720
	CACTTAAGAA CTTGCTTAT TAATTTTCTA CCAATACACT CCCTTCTAAG TGCACTAAAA	3780
	AATCCTTACT GCTAAGTGAT TAACTTAAC AATAAGGATT TATTTATCAT TAGTGGATGA	3840
40	TTATTAACGG AATCTCATAC CACCATCTAC AATAATTGTT TGTCCAGTAA TGTAATCAGA	3900
	GTCTTTACCA GCTAAGAAGC TCACTACATT TGAAACATCT TCTGGTTGAG AAACTCTGCC	3960
45	CAAAGCAATC TGAATTGTAA ATTGTTCCCA ACCCCATGCT TCAGGTTTAC CTGCTTCTTC	4020
	GGCTGTTGCC ACTGCGATAC TTTCCATCAT TGGTGTGTTGA ACGATACCAG GTGCGAATGC	4080
	ATTCACAGTA ATACCTTCAG ACGCTAAATC TTGTGCGGCT ACTTGTGTTA AACCTCGCAC	4140
50	TGCGAATTTT GTACTGCAAT ATAAAGACAA GCCTGGGTTA CCCTCAACGC CTGCTTGAGA	4200
	TGTTGCATTG ATAATTTTAC CGCCATGATT GAATTTTTTA AATTGTTTAT GTGCGGCTTG	4260
55	AATACCCCAT AGCACACCTG CAACGTTTAC GCCATATACT GTTTTAAACT GTTCTTCAGT	4320

	GCCAAATTGC	GCGGCAGTTT	GTCTTAcTGC	GTAAATACA	TCATCACGGT	TTGATACATC	4440
	TGCTTTGATA	GCAATAGCTT	TTGTACCATC	ACTTGATAAT	TTAAGTGCAG	CTGCTTTTGC	4500
5	CCCTTCTTCA	TTGAAATCAA	CAACTGCTAC	TTTGAAACCA	TCTTCCACTA	AACGTTCTGC	4560
	AATTTTAAAA	CCAATCCCTT	GTGcTCCGCC	AGTTACTAAT	GCTACTTTGT	TGTTTGTTCAT	4620
	AAAGATCACT	CCTCAAATTT	CTTTCCTTTA	ATTACATTTT	ACTCCTCTTC	ATTTGAATAG	4680
10	TACAACAAAG	GTAGCTCCAT	TTAACAAAAT	ATTCAGATAT	TTAAGGTATA	GTTAAACGCA	4740
	CTACCATTAG	TGATTGGCAA	TGCGTTTAAA	TGTCGTTTTA	AAAGTTCTTA	TGTTGAATAT	4800
15	TATTTTTTTT	AGTCTCTCGA	TTAGTTTGTC	ATCAATCTTT	TTTCGAGACA	TGGTCTTTTG	4860
	ATTCAATAGG	CGGTTCCGTG	TTATCACTGA	CAACTTTAGT	TGTAGCTTCA	TCTTTATGTA	4920
	TTTCTTCGTT	AAATCCTTCA	AGGTTTTTAG	TCGTGGGATT	TTTAACCTCA	GGATGTTCCA	4980
20	TCATGTCTTG	ACTATCAAGT	TCCTTTTTTAC	ACGTGTCTTT	ATGTGATGCT	TGATTTGCGT	5040
	TCCCTTTACT	TTTTTGAATA	GTGGTAGTAT	CTGCTGCAGC	TACTAATTTT	TTTCTACCTA	5100
	AAATAGATAT	GGCTGAAACA	AACCAGAGTA	TTGCAGATAC	AAAGTTGCAT	AATACTAAAG	5160
25	CGATAATAGC	CAATACAATT	AATATGACAC	CTTTTGAAAT	CCTTTCTTTA	AATAAGTCAG	5220
	ATGCCAATAC	GATGACAGGT	ACGATTGAAA	GTATAATTAC	AAATATAGAA	ATTATTGCCG	5280
30	ATATAACTAT	TGTTACTATT	AAATAATCAG	CTCTGCTACC	TGATAATAAA	TAGAAAAGGC	5340
	CGAAAATTAG	TCCATAGCAA	ATTACAAACC	CACATAAAGT	TATAGCCATG	AGTACTATAT	5400
	AAGCTATTTG	AAAATATAAA	CCTATCTTTA	TGAATGATTT	TTCTACATTT	TTTTCCATGT	5460
35	CTATTCCCCA	TTTATTTAAA	ATTTATACTT	TACCTTAAAT	ATTCTCTTTA	TTCTTTAGTG	5520
	ATTTTATCTT	TAGATTCAAA	TTGATTCTCT	GTACTTTCAA	TATCAACTTT	TTCATTTTCG	5580
	TCTGTTCGATT	CATCTTTTGA	GTATTTATTG	CAAATCAGCA	AAATACCACC	AATCAGCCAT	5640
40	AAAATTGACG	AAAGGAAATT	ATATAAACAC	AGTGCAATAA	TAGCATAAAC	AATAAAAAGT	5700
	GCACCTCCGA	TTACAGAGTA	ACTTTCCATA	TAAATCGCAG	TAAAGATGGT	TGGTAAAACA	5760
45	GTGAAAAGAG	CCAATATTAA	TCCTAATAAA	AAAATTGTTT	CGTAATCAGA	TCCTCCAGCA	5820
	ATATTAATAG	ATATCATCCT	AACAAAAACG	ACACTAAAAT	ATATTTGAGC	TACGATGCCT	5880
	ATCCAAATTG	CTATTTTTC	TATAATTGAG	CTCATACTCA	TTCCCCATTT	ATTTAAAATT	5940
50	TATACTTTAC	CTTAATATAC	CTTATTTTAT	TTAATTTTTA	TATGCAAAAT	ACAAAAATGG	6000
	AGAACTTCAA	TATTTATAAA	ATATCAAAAG	TTCTCCACAC	TATATTGTTT	TATTATATTT	6060
55	TCGCTATCAA	TACGCTAAAT	CATCATATTT	CCCTCAACAT	CACAGTAAAA	CTATTGCTCC	6120

	TTCCAATTGC GCAGTTGTTT AACATCATCA TCTTGTTTAA GTAATGCCAG TGGTACTTGA	6240
	AGATTAAAGAC ATCGTCCTGA AATATTAAAG CGTGTCCACAC CTGCTGGCAC AGTTTCCCCT	6300
5	TTATGAACAA CCGCTTCAAT TTCCTTATAA CTCAATGGCT GATACTTCAT GAGTACATCT	6360
	TGTTGAGAAA GACAAGGATA TGTACCTTGT GCAATTCTCT CTACAGAACA ACAACCACTA	6420
	TAACCTGCGA CAACCTTTTC CCATACTTGA AAATGTGCTT CGCCTAAATC TTTTGTATAC	6480
10	AAATATTGTT CTGTATCACC ATGACACATT GTAATAAATG GCGCTTCTTG TCTTGTCTCA	6540
	GTAGTCCATG GCAAGCGATG TTCTTGTGTG AACGTTTCCC ACCACACACC AAATGGAACT	6600
15	TTATGTGCCC ATGTACTAAT TGAATATTGT GTTTCATGGA TTTCTTGAC TGGAACTTTC	6660
	TTACATCCTA ACGCTTTCAA ACTTGTATAC CGATGCACAC CATCTATAAC CATATATCTA	6720
	CCATGTTGCA TCGCTGTCAC TAAAATAGGA TGACGTATAA AATCATCTGC TTCAATACTA	6780
20	CTTTTCGTTT TTTCCAATCT TAAAGGTTTG AATGTTTCGT GAAGATCAAT CTTATCTACT	6840
	GGTACCAATT TTAAATGTTT ATGAATATGA TTCAATAGTT ATTCATCCTC CTTTGTGTTG	6900
	GTTAAATAAA TAAATTCAGG ATGTGGATGG CTTAAGAAAT CGTGATGTGA AATAGACCAT	6960
25	CCGTATGCAC CTGCATATTT GAAAACAATA ACGTCGCCTG TACTGATTGC GTCTATCTGT	7020
	ACTTCTCTAG CAAAGACATC TTTCGGTGTA CATAATTGAC CGACTAACGT TGTGTCCTGT	7080
	CTCGAAATTG AAACTTTTTC AAATGAATAT GGATTGTCTT TATAGCGATA AATGTCAAAA	7140
30	GGATGGTTAT GTTGCCAAGA TACCGGCAGT CTAAATTGTT GCGTACCTCC TCTTAATATG	7200
	GCATACCAAG CACCATGTAC TTTCTTAATG TCTAGCACTT CTGTCACATA GTAACCAATA	7260
35	TGTGCCACAA TAAAGCGCCC ACATTCAAAG TTCAATGTCA CATCTTCCAT TTCTTGCTCA	7320
	ACGATAAGTG TTTTAAAACG TTCTACAAAA TTATCCCATT CAAATTGGTT AGTTAAATCT	7380
	GCATAGTTAA CGCCTATGCC ACCACCAAGA TTGATATGTT TGAGTGGAAA TCGATGTTTT	7440
40	TCAGACCATG CCTTTGCTTT TTTAAAATAA AGTTTCACTA CATCGACATG TAAATTCGAG	7500
	TCTAAATTGT TAGAAATAGA ATGAAAATGA AATCCATCTA GATGAATCTT TGGCATTGCG	7560
	AGCGCAGcTT cAATGACATC ATCAACTTCG TCTTCAGAAA TACCAAATTG TGTTGGGCGT	7620
45	CCTGCCATAT GCAACGTTGC ATTGGGAAAT GGTCCTGCTA AATTAACACG CAATAAAATG	7680
	TGTTGTGTCT TATCTTCATC TTCTAAGATG GCATTTAGCC GTTGTAATTC ATGCATACTT	7740
50	TCAACATGAA TACGCTGAAC ACCTTCACTT ACTGCATATC TTAGTTCCTC GTCTGTCTTA	7800
	CCAGGGCCAC CAAAAATAAT ATGATTTGCT GGTTTAAAAG CAAGACCTTT TGCTATTTCA	7860
55	CCTTGAGATG CAACTTCGAA TCCTTCAACA TACTGACTAA TTGTATCTAG GATTTTTCGT	7920

	TGTTGCAAAT GATGTTCCAG TCCGACTAAA TCATAGATAT AATGACAAAC TGGATGAGAT	8040
	TGTGCTTTTA ATTGTTCAAT AACAGGTTGA ACTATACGCA TTAGCCTTCA TCCCCTTTCT	8100
5	GTTTAGACGT CGCTAGAGAT GCACTTAAAT GGCGATATAT TTTCCGCGA TCATCACCTA	8160
	AAATAAATGT TTGTACACCT TGTGCCTGCC ATTTTGCAAT ATCTTCATCT TCACGTGGTA	8220
10	ATGCACAAAA ATGTTTACCA TGTGCATTCA CAACTTCAAA AATATGTTGA ACATGTGATG	8280
	TTACTTGATC ATCACGCGTT TGCCATGGTA TGCCAAGTGA CTGCGATAAA TCTGCGGCAC	8340
	CTTCGACTAT CATGTCTAAA CCTTCGACTT GTGCTATATC GTCAATGGCC ATAACCCCTT	8400
15	CAACATCTTC TATCATGGCA ATCACCATAA TATGCTCATT AGCCATCTCC ATTGCATCAA	8460
	GTAATGGTGT ACGTCCAAAT CTTGCCATGC GACCACCATT CAAACTTCTT AATCCTTGCG	8520
	GGTAATAACG ACTTAATTTT ACAATATGCT CAACTGTCTC ACGATCTTTA ACGTGTGGCA	8580
20	CAATAATACC TCTCGCACCC ATATCCAACA CTTTAATGAT ATCTCTATCT ATCACTGCAG	8640
	TGACACGTAC AATTGGTATA ATATGCGCTG CTTCAGCTGC ACGAATTAAA TGCGCTAGTG	8700
	TCTCATCATT AATCGCCACG TGTTCGTGTAT CAATCACAAC AAAGTCATAC CCGCTTGCTG	8760
25	CGATAACCTC GATCATCAAT GGGTCCGGTA TAGAATTAAA AATGCCATAA ACTGAATCAC	8820
	CATTGTTTAA TCTATGTTT AGAGATAGTT GTTGCATCAT TGATACCTCC TACACCTAAT	8880
30	GGATTTGTAA CATGATGAAT TCTTAACTCG GAGTCACTTA ATAATCGACG TGTCGTTAAC	8940
	TTTTCAACTT GAATCGTAGG TTCAAACAAA TCGAAATGTT GATAGTTATT CAACTCTGGA	9000
	AATGCTTCTT GATACGCCTC GATGATGCCT TTAACCCATT GCCATTGCAG CTCCTCATCG	9060
35	ATACCATATT GCTTTTCAAT AAATAAGATG ATTTCCGGCGA TATTAATAAA GAAAAATGCA	9120
	TCATGTAAAA AGTCGCGTAC TAAACGTTCC TCATCTGTTT CAATAAATGA ATTACTATTC	9180
	ACTTTTAT GTGCTTCTGG CATTGGCTTT AATGTCAGGT GTGAAGCAGC TTCACTTAAA	9240
40	TGCTCAGCT TAAAACGAAC ACCATCATGG AAATCTTTTA AGGCAATACG TGTAGGCCAA	9300
	CCATTTTCAT GAATGAGCAT CATATTTTGT GCATGCGATT CAAAGGCAAT ACCGTGATAA	9360
45	TAAAGCATAT GAATCATTGG ACGAATCGCT ACAGCTAAAA ATTGCTTTGT CCAAGCTTCA	9420
	GAACCATATT GTTTAATCCA ATTTTCAATG AATGGTACAC CATCCTTATC ACTTGCATAA	9480
	AGTGCATTAA ATGGTATCGC ATCCTCTTCA TCGATTAAAC TATGATATAT ATTTTCACGC	9540
50	CATATAACAC CTAACGCACC ATAACTTGA GTTTGTTTAT AAGGCGAAAG TTGTGTATTT	9600
	AAATAAGACT GTCCTAAGAC TTCCCCTAGA AAAACTGTCT TTAATTCATC TTTTAAATAC	9660
55	ATATCTTGTT GCTGTATCTG CTTTAACCAA TCCGTAATTT GCGCTGCATT TTCAATTGTA	9720

	TATTTTGTGCG TGTCTATTGG CGACATCGTA CGAATCGATT GTTGAGGGTG ATATAGCTCA	9840
	TCACTTTCCC CTAACCATAG TACTGTGCCA TTAAGCCTTT CTTAGCCAA ATCAACTTGG	9900
5	ATGACATGTT CAAACTGCCA TGGGTGTACA GGTATCATCT CAACATCATT TACATGTTTG	9960
	CCAGATGCTT CAATTTGCTG TACAAAATGT TCATAAGTCT TATCGCCAAC TTGTTGACGT	10020
	AACATTTCTG TAACTACAAC ATTTCTTGAT ACCGTCGTTT CTACTTTATC TTTGTCGATA	10080
10	GCTAACCCTT GCAGTTTAAC GTTTGGTACA AAATCAGGAC CAAATTTCAA ATTATCACTC	10140
	AACGTAAATC CTAAACGTGA TTTGTAACTT GGATGATACT GATGCCCTTC CATCGCATAA	10200
15	AATTCATAGT CGTTAAATGT CTCAGGTGTT GCTGGTGGGT TTGATTCTCG ATACTGCATA	10260
	CTTTGCGTAT CTTTTAATTC TGTCTGTAAT AACTCGACAA TAAATTGTTT TAGCTTTTCA	10320
	TCATTTTTAG GAAATGTAAA TACAACCTCT CTCAATAATT GTGTATAGTC TGTGTTGTA	10380
20	TCTGCCTCAT CTCCTACGAC ACGCTCAATT GGTGATGTGA TACGTATACG ATCAAAGCTA	10440
	TGTGTCTTTT CAGCAGTAAA ACGATACTCT GAATCATGTC CTTCTATTGT AAAATGACCG	10500
	ACACCGTCTT GATATGACGC TTTATACACA ACAATATTCT CATAAATAAG TGATGATACC	10560
25	AGTTGGTGCA TCACTCTAGT CTTTACACGA TTAAGAATTG TTTGATTAC AATACGATAC	10620
	CTCCTTGTTA TGACAAATTG GATTTGGTAT ATGTGTATAA ATAGGGTTTG CACCACAATC	10680
	ATTCAATTTA CTCATCAAAT TCGCTTTAGC CGCAATGGTC GCGTTTGAT ATAAATCTTC	10740
30	TACACAGTCA ACAAATACTG CGTTATTTCG GTATTCTTTT TTCCAAGTCA TAAGACGATG	10800
	CGCTACAAGT TGCCATAACA CAACTTCATT TCTAGTCGCT TTACCAATAG TTGATACTAA	10860
35	ATGTCCTAAG TGATTTACTA CAACGTAATA TTAAAGACGA TGCCATGCTT CATCATGTGC	10920
	ATATACAACA GGGCTTGATG CTGCCACAAC ATTTGGCACA AGCTGTTTTT CAGTAGCAAT	10980
	CGTTCTAGAT AGACAAATGC CTTCAAGATC TCTGACAAAG CATACGTCGG GTATGCCATC	11040
40	TTTTAATTC AATTAATGTAT TTTGTACATG TGCTTCTAGA CTAATGCCTG TGTTACTAAA	11100
	CAGCTTTAAT ATCGGCAATA ATGTACGATT CAAATAACAT TCAAGCCATG CTTCTGGTGC	11160
	TAAACCACTT TGCTCAATCA CTTGTGATAA CTTAGACATC GGTGAATCAG GCATCGTTTC	11220
45	AAATAATGAC GCCAATACAT GAATATCTTT ATCAGCATGG TAATTCGGTA TCCCTTCACG	11280
	AACAATCATG GCACTATTTG TTAATAAATC CATTTCAGGT TCAACTGTTT GCCCTAATGG	11340
50	ATTCGGTAAC AATGCACGAT ATCCTTCTTC AAACATCAAT TAAAAATGGG GTGTTTCAAC	11400
	CTCATCTTTG ACTGATGCGA TAACTTGCGC GGCATCAATT GTCCGTTCAA TCTGTTCAAG	11460
	GTCATTCGTA CGTATAAAAT TAGTGATTTT AACGTGTATC GGTAATTTTA AATAAATGTT	11520
55		



EP 0 786 519 A2

	GCCAAGGTCT	TTTATTAAAC	CTTGTTCACT	ATATTGCATA	TACTGTGGAT	GCTGTGCGAA	11640
	CACATTGATT	TGATAAGGAT	GTGTTGGTAA	TAAAATAAAA	TCTTTGGGTA	TCTCTGATAT	11700
5	ATCTATGTCT	GCTAATTGAT	ACAACACTTT	CTCAACCTGA	TCTTCTTTAC	CTTCTACATA	11760
	GCGCGTGAGC	AGAACATCTT	GATGCACAGC	TAAATAATGC	AATTGGAATG	ATGTATGACA	11820
10	TTCGGGTGCA	TATTTCTCTA	AATCTGCTTC	TGAAAACCCA	CTTGCACTCT	TAGGAGTCGG	11880
	ATGAAATGGA	TGACCTAAGT	ATAAAGATTG	TTCTGAAACG	ATATAACGAT	CCTCTACGTA	11940
	GTCTATTGTG	TTACTTTGCA	AATAACGTGC	CGTGCGATGA	ATGCTATTAT	CGATGTCAGA	12000
15	CATAATTTGC	GCCATATGTT	GTTGCACTGC	CGTTTGATTA	TCTGCACTTT	GAGCCATATG	12060
	TTGCAAAATA	CGCGCAATTG	CTTCTTTATA	AGTTGTTATT	TTTTTACTTT	TTCCATCGAT	12120
	AAGCCATACC	TCTGGATGAT	ACATATGATG	CCCCATCGCA	GACCAATAGC	GAAATTCACC	12180
20	CGTTAAAGTT	TCGAGCTCTG	ATAATTGTAT	AGACCATTGA	TGATTTTGAG	GTGGTACTTG	12240
	ATATAAATTT	TCTTCTCTAA	AATATTCATT	TAAATGCGT	TCGATAGCCG	CATACGCTGC	12300
	ATGTTGTATT	AATTCCTTAT	TTTGCACTTT	TTTGTTCCTA	CTCCATAAT	TTCATTAATG	12360
25	TGTGATCGTT	GATTTGATTA	GTGATGGTTG	AACAAATTAA	AAATAAACTA	CTTACTGCAA	12420
	ATACTACGCC	CATAACGATA	AACGTAGTAG	CTGGTGTAGT	ATAACTTGTA	ATGGCAGCGC	12480
30	cACTaAGACT	GCCAATAATT	TGACCAACAA	CTAACATACT	GTTCGTCGTT	CCAACAAATG	12540
	TGCCTTTAAG	TTGTTGATGA	CACGCATTCA	CGACAACAAA	CATGACACTT	TGAATCAATG	12600
	CACTATATGT	TAATCCTTGA	AGTATTCTTG	CAGCCATTAA	AAACTCTATA	TTCGTCGCTA	12660
35	AACCTTGCAG	TATCGCACTA	CAACCACATG	CAATCGTGGC	AAATATATAT	ACTGATTTAA	12720
	CATATGATTT	ATCATTAAG	CGTCCCCATA	AAGGCGCGCT	TAATATCGAA	GCCGTCCAAA	12780
	ATGCGGACTG	TAAAAATCCA	ATCACACTAC	GGTCATCTAT	CGCTGTATGA	TTCACTGATG	12840
40	AAGCAAGTGG	TGATAATGCA	GTTAGCATGC	CATACATAGC	AAAGTTTGCT	AAAACGCCAA	12900
	CGATAATAAA	TCGACATGTT	TGTTGTGTGC	ATAATAGACA	TTGAAATGAA	CGGCGAATAC	12960
45	CTTTATTAAT	ATTTGGTGTT	TGTGATTTTG	GCATATGTGT	CGTTTCAATC	AATTTTAATG	13020
	CACCGAAAAT	ACAGACAATA	AAAGTAATAA	CGGCAATACT	CATCAGTAAC	GCACTAAAAC	13080
	CTAATATCGA	AGCTGTAACA	CCGCCAATTA	ATGGCCCCAC	AAGAGACCCT	GCGCTGACTG	13140
50	AACTTTGCAG	TCTTCCTAAT	ACCTTTCCAC	GATCTTCAGC	TGGCGCCTCT	GCACTCGCAA	13200
	ACGCACTTGA	TGCATCAACA	ACACCACCAA	ATAGTCCCTG	CAATAACCTC	ACAAGTACAA	13260
55	ACTGTAATGG	TGTCGTACAC	AATGCCATTA	AAAATAAGCA	TACCGCCAAA	CCAAGTAACG	13320

	ctATCATCGT CGTTACAGCT GGAGCAGCAA TCGCTATACC ACTCCACAAC TGTATTTCTA	13440
5	CGACTGATAG ATTTTGTAGT GATGCCATAT AAATTGGCAA TAATGGCACA AGTACTGTCA	13500
	GTCCAGCAAT CGCTATAAAC TGA CTGAGCC ATAAAAATGCG AAAGTTACTG CGCCATATAG	13560
	ACTGATTAAT CATATGTCAC CATTGGATTT GGTACGGTAG TTAAACCTGA AGGCATACTA	13620
10	CCTCCACCAC TATCACGTTG ATATAGCAAT GGTAATAAAA TTTGTTTGAA TGGCCACGTC	13680
	TGTTTATCAA ATAAAAATGTG TCTGACAGCT AGCTGATCAG TTGTAACCCA GGAAATAGTT	13740
	GCCACTTCAT TTTTAAAAAT TTGTTTAAAC AACGACATAA GTTCATGCTC ACTTACACCA	13800
15	AATAAATCTT GAATTGCATC AATAATGGCA TATAGATTTA CCGATACAGC TAATGTTTGA	13860
	AAATAAGCAA AGAATGTTTC CAAATCCTCA TTAATTAGCG TATTAGGTGT ATCTTCTCTG	13920
	ACGACATACT TCGGCAATGA AAGCTGATGT GCTGTTAGCC ATGGTTTATA AATTCTGACA	13980
20	GTATCATGAT CACGTAACAC GCATTTTGTG ACACGTCCAT CTTCAAATGA CAACAATATA	14040
	TTTTGACCAT GCAACTCTGG TAATGCGCCG TATTGCATAA ATGATAGTGT TACCTTTAAA	14100
25	AAGACTTGCG CGATATCTTC AAATAACGTC ATGACATCAT TTTTAGAAAT ATTATCTTTT	14160
	CCACAAATCA TTTGATATAA AGTGCGATCA TTTGCCGCGA GTGCTGCCAT TGACACTAGC	14220
	TGTTGCGTAT CATTTTGGC TAGCACTTCG GGATACTTTC TTAGCTGAAC AGTTAGATGA	14280
30	CCTAATTGAT CTTTGAAAAT ATCATTATCT TGACCCATAT ATGACCACCA AGCTGTTTCA	14340
	TCACAAACCA TGACATACTT AGCTAGTGCT TCATCTTTTT CTATAAGCTG ACGTAATAAT	14400
	TGTTCTGCTT GTTCTCCGTT TTTTCATGTAA CGCGTAGGCG TTAGCCTTAA TGCGCCTAAT	14460
35	GACTGCATTG CAAATGGTAC TTTGACATGG TTATACGGTG CGCCAATATC AATTAATGAA	14520
	CGCATACTTG AAGACGACAG ATAATCTCCA AATTTTAAACG GTAATAGTAC AACCAACTTT	14580
	TCAC TAATCT CTTTCGCAAA GACGTTTCGGC AGAATATGCT GATATTGCCA AGGATGTACC	14640
40	GGAAATAGTA CATAGTCATC TATTGATAAC CCTTGATCAT TTAACATGTC TGTCGCTTGT	14700
	TCTTTTATAG GTACTGTCAA ATTTTCTAAT TCATCGATAT TTGCAGTATC GCCATGAATC	14760
45	ATATGTGTCT TTTTAACTGC TGCAACCATT AAAGGAAATG ATTGATTTAA TTCAGCTTGA	14820
	TACACTTGAT AATCCGCTTC TCTTAATCCT CTTTTTCTT TAGCTAATGG ATGAAATGGA	14880
	CGATCTTTTA AACTTGCAAA CTGCTCTGAC ATCACAAAAG GATGTGACGC TAAATCTAAT	14940
50	TCTGATAATT GTTTAGCAAG CTGTGTGGCA GCAGTAGTCA GTCCTTCTTC AACGCGAGCC	15000
	ACTTCCCATT CATGACTTAG ATCACAATTC ATATTAGCAA TTGTTTGCCA AAATTCAGCT	15060
55	GCCGTTAAAG GTTGCTTAGA CACCCTTCCC TCTATCGTAA TTGGTTGTGA ACTTTCGTAA	15120

	TATATCAAAA GCGTTTGTCC GTTTTCTTTA GTAATCTCAC TATTCGATAC AATTCCGGCT	15240
5	ATATCTTCAA ATAATAATGC ATCAACTAAA TCTCTTAATA TTATCGCTTG TGCTGTATTG	15300
	ACTGCTGTAT GATTCTGCAA TGTTCTAGACA CCTCGCATTG TTAATATAGG TTCAATGTTG	15360
	TCCCAATATT TTGTTGTTGT GCCTGTTGAT AAATAAAATA AGCACTTGAA ATATCTTCTGA	15420
10	TAGCCATACC CATCGGATTA AGTAATATGA TCTCATCATC GTCTTCACGT CCTGGTATGT	15480
	CACCTGTCAC AAGTTGTCCT AGTTCAGCAT GAAGAGCTTC TTTGCTGAAT TTACCTTCTA	15540
	ACACCAATTG GTTAATAGTT TTCTTTTCTC GATTACATTG TGACCAGTCA TCTACTACGA	15600
15	CTTTGTCAGC TTTAATAAAG ACTTCTTTAT GCACATCCAT GATAGAAATG TTGCTAATAA	15660
	ATGCACCCTT TTGTAACCAA TCATATTCAA TGTATGGTTG ATCCGTTACG GTACATGTAA	15720
	TGACTACTTC ACCATTTGAT ACTGCTTCTT TAGCATTTTC TGTCGCAATA AAATTAATTT	15780
20	CCGGACGCTG TTGTTGCCAT CTATCAACAA AGCGTGCACA TGCTTCAGAG AATTGATCGT	15840
	AAACAAACAC GCGTTCAATA TGATCGAATT GCTCTAACAT ACTTTGTAAT TGCTTGCTCTC	15900
25	CGATTAGCCC GCATCCAATG ATTGTTAAGT CTTTAAATCC TTTTTTAGCC AAATGCTTTG	15960
	CTGCAATCAC TGAAACTGCT GCAGTACGCA TACTACTAAT TAAACTTGCT TCCATAACTG	16020
	CAATTGGATA ATTCGTTTCT GGATCATTCA AAATAATGAC GCCACTTGCA CGCTCCATAT	16080
30	TACGTTTCGA TGGATTGTCG TGCTTACTAC CTATCCACTT AATACCTGAA ATTGCGTGTT	16140
	CACCACCGAT ATGACTTGGC ATTGCAATAA TTCGATCTGC GATGTGTCCA TTTTCAGGAT	16200
	CCTGTCTTAA ATACGGCTTA AGCGGTTGTA CAAAATCATT GTGCGCATGG GCTGTTAATG	16260
35	CTTCTGTTAA TGCGTCCACA TAAACTTG TG AATGATTACC TCCCGCTTGT TCAATATCTG	16320
	ATCTATTTAA ATACAACATC TCTCTatTCa TTCTGaTTTA ACTCCTTGTC TTGATTTTCAT	16380
	TTTTTCTAAC CATGTATCTG AATAAACTAA ATCTAAGTAA CGATCGCCTC GATCTGGTAA	16440
40	AATCGTGACA ATTGTTGCAC CTTCTTCAAT TGACGTTATC AACTGCTCAA TCGCTGCAAT	16500
	AATCGAACCT GTTGAAcCTC CGGCAAATAT GCCTTCATAA TCAATCAGTT TTCGACAGCC	16560
45	CAAAGCAGAT TGATAATCAT CTACATGGAT CACTTGATTA ATTTCTGATC TATTCAATAT	16620
	TTCGGGTACA CGACTAGCAC CGATACCAGG TAATTCTCTA TTAATAGGTT TGTCACCAAA	16680
	AATGACTGAC CCTTTCGCAT CAACAGCAAC AATTGTGCG TTTGGATGCA CTTCTTTTAT	16740
50	TTTTCTACTC ATACCCATAA TGCTACCTGT CGTGCTGACT GGCGCGACAA AATAATCTAT	16800
	AGGTGCTTA ATTGTTTCAA CAATCTCTGT GCCTGCACCA TGATAATGGG ATTGCCAATT	16860
55	TAACTCATTG GCATATTGAT TAATCCAATA TGCATCGTCA ATAGTGGCTA ACAGTTCTTG	16920

TACATTGGCA CCATAACTTT TAATAATTTT CAAATTTGTT GGTGATATTT TAGGATCAAC 17040  
 AACACACGTG AGTTTTAATC CCTTGATTTT AGCTATCATT GCCAACGCAA TGCCTAAATT 17100  
 5 ACCAGAAGTA CTTTCAATTA AATGTGTATT CTCAGTGATT AAACCATGTT TAATACCATG 17160  
 TTCAATGATG TACTTGGCAG GTCGATCTTT CATGCTGCCT CCAGGATTCA TATACTCTAA 17220  
 10 CTTTGCAAAC ACTTCATGTT TCGGAAATAG TTGATGAAGT TGAACCATAG GTGTTTGCCC 17280  
 TACAGAATCT AACAAATGAAT CGTGACATG 17310

## (2) INFORMATION FOR SEQ ID NO: 24:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5423 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

25 ATACTAGTAA GCGCATCGGT TATTGACATC GAATTCAACT TTAACAGTTT TCATGTTTCGG 60  
 TGATGTTTCa ATAGAATGTG TGTGTTGTAC TTGCGCATTT ATATTTCCAC CTAAATTACT 120  
 TAAGTTTCCT GTAATACTAG AAATGTCAGG TCGGTTTAAT GTAGGTTGAA ATGCATCAAC 180  
 30 TACTTTATCT GCAACATTAG AAACATTACG GATAACTTTA CTTGAATGAT TATCTATACC 240  
 TTTAACGAAA CTAACATTG AATACATACC AACATCCATG AATTCACGTG AAGGTGAGTG 300  
 AATACCTAGC GCTCTTTTGG CTGCATTTAA AGCACCTTTT GCTACACTAG CTGCTTTTTTC 360  
 35 AGCTAAGTCT CTAGCCATAT TACCAATACC TCTCATCAA CCACGGATCA TATCAGCACC 420  
 TGCTGATACA AAGTCATCCA CAAAGCTTTT AACTTTATTT ACTGCATTTG TCATACCTTG 480  
 ACTA~~A~~CTTTG TTTACAACAT TAACGAATCC TTGAACAACT CTATTAACAA rGTTAATTAG 540  
 40 CGTACTtGTt ATAGTAGATA CCCaTnGCAT ACCTTTAGTG ACmATGAAGT TCCAAGCTTG 600  
 AGACATTTTG TCTGATATAG TTGAAACAAC TTGTGTGAAT ATGCTTACAA CTTTATTCCA 660  
 AATGTGCGTT AATATACCAG ATAAGAAACT CCAAATCGTA TTCCATATAT TAGAAATAAA 720  
 45 ACTCCATGCC GCTTGTAACG CAGTAGATAT AGCTGTAGTG ATAGCGTTCC AAACCTTAGT 780  
 TGCCACAGTA ACTATAGTGT TCCACAACGT TTGTAAGAAC GTCCAAATAG CGTTCCAAAT 840  
 50 TGTTATTGCG ATAGTCATAA TTGTGGTAAA CACTGTAGTT ATTACAGTGA CTAACAAATT 900  
 CCAAATCGTA GTAGCGATTG TAATTATCGT ATTCCAGATT GTACTTAAGA ACGTCCAAAT 960  
 AGCTGTCCAT ATCGTCATAA CTATTGTCAT TATCGTCGTG AAAACAGTTG TAATGATTGT 1020

55

	ATAAGCGACT ATTTGATTCC AAACAATCAT TATAAAATTG TAAACATTCTG ATACTGCTGT	1140
5	AGTGATAGCT GTTAAAATAG CATTCCATAC AACC GAAGCT ACAGCTTTTA ATACATTCCA	1200
	AACATTAAACC ATAAACGTTT TTATCGCATT CCAAGCATTT ATAATAAAGT TTCTGAATCC	1260
	TTCATTTTTTA TTCCACAATA AAACGAATAT AGCTATTAAT GCAGCAATTA CACCAATTAC	1320
10	TATTGTTATT GGACCGCCTA AAATACCAA CACAGTTACT AGTCCTGTGA TAGCATTCT	1380
	AATTAATCCA ATCTTACCGA ATAACAATTG GAATATAACT GATATAATTT TTAATGGTCC	1440
	TTTTAATAAC ATGAACGCAC CTTTTAAAAT TGTTAATCCC GCTCTTAATA AACC GAACTT	1500
15	ACTTACTAAT GCAATGTTTC TACCTATTAA TCCGCCACCC ATAAAGTTAG ATACAGCAAG	1560
	AATAATCGGT ATTAAAAATC TAAATGCACC AACTAAAGTT ATAATGACAC CAACTAATTG	1620
	TGCTGTAGCT GGATGCGCCT CAAACAAGTT AGCTATCCAA CCAGTTATTG CAACTGCAAC	1680
20	GCGTAATACT GCACTAGCTA TAGGAGCCAT TGCTGTTGCG AATGCATTA ATCCTCTTGC	1740
	GATGTTTCCA ATCAATTGCA TTATTAGTGG TCCATTTGTT TGTATATAAC TGACAAAGTC	1800
25	TTTAAACCCT TGAGATTGTC CTAATTGTTT AGACCATTC CTAACCTTAG CTGTCATTTG	1860
	TTCAAGAGAT TGGAATATGC CAGTTGATGA TCCGCTGAAT GCATTCATCA AATTGTTAAT	1920
	TCCAACGAAA ACATTTTTGA AAATATTACC AATGATAGGT AAGTTTGTTT TTGTGTATT	1980
30	AATAAACGA GTTATCGAAT TTTCTCCAGC TGCACTATTA GCCCAGTTAG AGAAAGATTG	2040
	ACCTAATCTA TCCAACCAAT CAGCCGACCA TTGAAACAGT GGTGCTAATT GCGTGAATAC	2100
	ATTGACTAAT CCGTCACCAA AACCACCTGC AGCACTTAAT AGCTTGTTAA ATACCGAAAC	2160
35	ACCGTTGTA TTCATCATAT TAAAGAATCT TGAAGCTACA CTGCTATTTT CAGCCCATT	2220
	AAGCACGCTT TGAGACGCTT CTTCCATTCC TCTTGAAATA CCACTAAAAA ACGGTTGTAA	2280
40	GCTCTGCATT GCAGTTTTAA CAGTATTTAA ACCATTGCA AGAGTTGTGA AGATAGCGGA	2340
	TTGATTTTGC TTTATAATAT CAGTCCATGC TGACTTTACG CCATCTAACG CTTTTTTGTA	2400
	TTCGTTTGTT GCTGAGCTAG CTTGTAAAGT GCCATCATT AGCATCTTTA TAGCGCTGAT	2460
45	AGCCATTGCG CCAAACGCTA CAAATCCTGC TCCCGCTATT GCTACGGCAC CACCTAAAGC	2520
	AAGTACACCA CCAGTTAACA CTTTGATAGC GTTTAATAGC GCAAATACTA CAGGTACTAC	2580
	GCTCGCTATT ACAGGTATTA AGATACTAAA AGATGATGTA AGTAATCCAC CAACCATATT	2640
50	AGAACCTACA GTACCGAACA CACGGAACAT ATTAGCTAAA TTCCCCATCT GTCTTTGAAA	2700
	ATTGTCATTT GCTTTTATTA TGTAGGCATA AGCTTTCTTT AAACCATTAG TATCGACATC	2760
55	TACCTTTGTT GTTTTTTTGT TCGGCAATGC GTCTAATGAT TTTTAAACG CATAAATAGT	2820

	AAGTTCTTCT	TTAGTACGTT	TGATTTTAGA	GTTAGCAACA	CCATTGTCCA	CGTCTATAAT	2940
	AGCTTTGGCT	TTAGACCTAT	TTAATGCTTC	GAGACTAGCT	TTAGATACTT	TTAACAACCTCG	3000
5	ATTGAATTTA	CTGTTATCTG	CATTGACGTC	AATATTGACA	CGTTTCTTTT	CTAATTCTGA	3060
	TAATTTAGCT	TCTGTTTCAG	CGATATCTTT	AATCAACTTT	TGTTTTTGCA	ACTTAACTTC	3120
	TGGTGTAAC	TCTTTAGAGT	TTAGTTTGTC	TAGTTCAAAA	TTCGATTCTA	GTACCTTTTG	3180
10	TTGTAAATCT	TGTATACTAG	CATCTAATTT	AGCTTTTACA	TTTTTGTTAC	TAAAGGCATC	3240
	TAAAGACTTT	TTAGCAACTT	TGATAGTTTT	TTGTAAATTT	TTATCGTTAG	CGTTTAAATTC	3300
15	AACATCTTTA	GTTTGATCTG	CTACTCGTTT	AAATCTTTGC	ACAGACTTAA	CCGCACTATC	3360
	AATTTGCCCT	TTGAATTTGG	CTACACTAGC	TTCAATAGTC	GCTTTAATTT	TATATTCCGT	3420
	CACATTAAAC	CCTCTCTTTC	TATTGCTTAT	TAAATTCTGC	TATAACTTTA	AAGAATTCAT	3480
20	TATTTTGTGG	TTCGTATTCA	TCACGTTTCG	TACTAAATCT	TATATCTTTA	CCTTCGTTAA	3540
	GCCGTTGGAT	ATTTTCTTCA	TAAGGCAATA	CGTCGTTTGC	ATTGTTAAAA	ACATATTCCT	3600
	CTTTAGGTTT	ATTTTCTGTC	CCAACATTTT	TAGTAGCTGC	AGCATCACGA	ATAGCAAACG	3660
25	CAAGTTTGTA	ACGTTTGAAT	TCTTGGGTTA	GCATTTTATA	CTCTTTTCGA	TACATTCGAT	3720
	AGTTATATTC	TGTTAATGTC	ATTTGCTCAA	TAACGTTCAA	ATCTGTAATA	CCAAGTGTG	3780
	ACATACAAGT	TATAACGATT	CTGTCGTAAG	TTATTAGGCT	TCCGCTGGTT	TTTCTTCCGT	3840
30	TTCCACTACT	TCGACTAGGT	TTCCGGTTCAT	AGGTCGCTTT	CCCAACTCCG	TTAAATATATC	3900
	CGAACCGAAT	TCTTCTAGTC	CGATATTTTC	TGCGATTTCA	TCTAATGCTT	CATCAATGTT	3960
35	ATTAATAGTA	ATTGCTTGTT	TTTTTAAGTG	AGATGTAGCT	GCGATTAAAA	CTTCGCCAAT	4020
	CACAACCGGA	TTTCCACTTT	CTAAACCTAC	AGGCAACATT	GATACACCTT	GACCGATAGA	4080
	AGCTTGGTTCA	ACTTTTAAAC	CTAATCGGTT	ATCGATTTCT	CTTAAAAATT	TAAAACCAAA	4140
40	ACTTAATTCT	AATGACTTTC	CGTTAATTTT	TACATTCATA	ACTTAAATC	TCCATTCATA	4200
	ATTAATTTAA	ACAAAATAAA	TAAGCTTAAC	GCCCTATTTT	TATACCTCTC	TTGGTGCAAC	4260
	CGGTGGTGAA	TCTACTTTAG	GTTGTGGAAT	TGCTGTTAAA	TCTTCGCCAG	TTAATGCATC	4320
45	TGCTTTTGTA	GTGTCGTGGA	ATCTGTATCT	AGTCGCCTTA	AGTTTCTTTG	TTACAGCCTC	4380
	AGGTAGTGTT	GCAAATCCAC	GTTGGAAACG	ACCATTCACT	CCATATTCAT	ATTCATATTC	4440
50	ATCAATACCG	TTAGCTTCTG	CTTTTAATTC	AAATTTATTG	TGGAAACCTT	GGAAATATTT	4500
	CGCTTTAAAT	TTAGCGGAAT	CCCCATTTT	GCCTGGTATT	CTACTTTCAA	CTTCCCAAGC	4560
	TTCATACAAT	ACGCGATCTA	CAACTGCATC	TTCAATTTCA	TCTGCAAAAT	CGTCACCATA	4620
55							

GTCCATTGTA TCCTCTGTAT CTGTATCAGC TTCATGTGAT AAGCCGTATT CAGTTAAAAA 4740  
 AAGCATTTTA GTAGCATCTA CTTTTTCGCC AGCTTTTCTA AATAAAATAA TACGATCATT 4800  
 5 ACTATTTTTC ATATTTGCCA TTCAATATTC CTCCGTTTTT TAAAATGTTT TGTAAGATAT 4860  
 CGTTACTGAT GTGTGTAGCA ATTCTTGATT GGTAGTATCA TCAACTAACT GTGTGATGTT 4920  
 10 AGTATCTTCT TCTTCAAAGT CATAATCGTT TGTTTTAACG CTAGGTGTTA AATCATCAAT 4980  
 ACATCTTTTA ACAAGTCCGT CATGATGTCC TAAATCATCG CTTACACTCC AAATATCAAT 5040  
 AACTAAATTC GTATCGCCAG AATAACTATC AAACGTGTAC TTACTTCTAT TTGACTCCGG 5100  
 15 CATTTTTATT ACAAAAAAAG GATACGGAAT CTCTTGTGTC ATCTCTTTAC GAGAAATAAC 5160  
 AGGGAATCCA TATCCTTGTA GCGTTTCATA CGCTTTATTA TAAAGTTGTA AGTTCGGTGT 5220  
 CATGCTTTTA TCTCCTATTC AAACAACGCT TTCAATTCTT CTACAGTTGA TTTCTAATC 5280  
 20 ACTTCGTATA CCGGCCACAT AAAAGGTTCA GCCTCCATGT ATCGAGTACC AAATTCTAAG 5340  
 AAACCACTAT AAGCTGCGTG CGATGTGATA GTGTATTGCA AATCGCCAGT TTTTPTATAT 5400  
 CTGATATTGC GTGATaAATT ACC 5423  
 25

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6251 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AAACGCAGAT GTTCAATTAG AACCAGTCTA TCGTATTAAG GAAGGTATTA AACAAAAGCA 60  
 AATAAGAGAC CAAATTAGAC AAGCGTTAAA TGATGTGACA ATTCATGAAT GGTTAACTGA 120  
 40 TGAACCTAAGA GAAAAATATA AATTAGAGAC CTTGGACTTT ACTTTGAACA CATTACATCA 180  
 TCCTAAAAGT AAAGAGGATT TATTACGTGC TCGTAGAACC TATGCATTTA CTGAACTGTT 240  
 TTTATTGAA TTACGTATGC AATGGCTAAA TAGATTAGAA AAGTCATCTG ACGAAGCAAT 300  
 45 TGAAATTGAT TATGACATAG ACCAAGTTAA ATCATTTATT GATCGTTTAC CTTTGAACCT 360  
 AACTGAAGCA CAGAAATCCA GTGTTAATGA AATTTTTAGA GATTTAAAAG CACCAATACG 420  
 50 TATGCATCGA TTAATTCAAG GTGATGTAGG TTCAGGAAAA ACAGTAGTTG CTGCAATTTG 480  
 TATGTATGCG TTAATAACTG CTGGTTATCA ATCAGCATTG ATGGTACCAA CTGAAATTTT 540  
 AGCAGAGCAA CATGCTGAAA GTTTAATGGC TTTATTTGGA GATTCTATGA ACGTTGCATT 600

55

	TACGATTGAT TGTTTAATTG GAACCCATGC TTTGATTCAA GATGATGTGA TTTTCCATAA	720
	TGTTGGTTTA GTAATTACAG ATGAACAACA TCGATTTGGT GTGAATCAAC GCCAGCTTTT	780
5	AAGAGAAAAA GGTGCAATGA CGAATGTGTT ATTTATGACA GCAACGCCGA TACCAAGAAC	840
	ACTAGCAATA TCAGTTTTTG GTGAGATGGA TGTGTCTTCA ATTAAACAAT TACCAAAAGG	900
10	TCGTAAACCT ATCATTACTA CTTGGGCAAA GCATGAGCAA TACGATAAAG TTTTGATGCA	960
	AATGACCTCA GAGTTGAAAA AAGGTCGTCA AGCATATGTC ATTTGCCCGC TAATAGAAAG	1020
	TTCTGAGCAT CTCGAAGATG TTCAAAATGT TGTGCGATTG TACGAGTCTT TACAACAGTA	1080
15	TTATGGTGTT TCCCGTGTAG GGTATTGCA TGGTAAGTTA TCTGCCGATG AAAAAGATGA	1140
	GGTCATGCAA AAGTTTAGTA ATCATGAGAT AAATGTTTTA GTTCTACTA CTGTTGTTGA	1200
	AGTAGGTGTT AATGTACCGA ATGCAACTTT TATGATGATT TATGATGCGG ATCGCTTTGG	1260
20	ATTATCAACT TTACATCAGT TACGCGGTCG TGTAGGTAGA AGTGACCAGC AAAGTTACTG	1320
	TGTTTTAATT GCATCCCCTA AAACAGAAAC AGGAATTGAA AGAATGACAA TTATGACACA	1380
25	AACAACGGAT GGATTTGAAT TGAGTGAACG AGACTTAGAA ATGCGTGGTC CTGGAGATTT	1440
	CTTGGTGTT AAACAAAGTG GaTTGCCAGA TTTCTTAGTT GCCAATTTAG TTGAAGATTA	1500
	TCGTATGTTA GAAGTTGCTC GTGATGAAGC AGCTGAACTT ATTCAATCTG GCGTATTCTT	1560
30	TGAAAAATACG TATCAACATT TACGTCATTT TGTGAAGAA AATTTATTAC ATCGTAGTTT	1620
	TGACTAATTG CCATGCTGAT TTGTCAATTT GAGTGCAACa CTTGTTAAT TGAGTGATAT	1680
	GACACTTGAA CTATTTAAAT GTAAAGTGGT ATTTTAACAA TTTATAAATT TTCGACTAAA	1740
35	TAATAGCTAA ATATTACAGT TATTTGTTGA GTCGGTTAAA TAGAAAGTGT TATGATATGT	1800
	GAGGAATGTT TAAGACTAGG TACTAAAAAA TGAGGGGTGA GACGTTGAAA CTAAAGAAAG	1860
	ATAA <sup>~</sup> ACGTAG AGAAGCAATC AGACAACAAA TTGATAGCAA TCCCTTCATC ACAGACCATG	1920
40	AACTAAGCGA CTTATTTCAA GTGAGTATAC AAACAATTCTG TTtAGaTCGC ACTTATTTAA	1980
	ACATACCAGA ATTAAGGAAG CGTATTAAAT TAGTTGCTGA AAAGAATTAT GACCAAATAA	2040
45	GTTCTATTGA AGAACAAGAA TTTATTGGTG ATTIGATTCA AGTCAATCCa AATGTTAAAG	2100
	CGCAATCAAT TTTAGATATT ACATCGGATT CTGTTTTTCA TAA <sup>~</sup> ACTGGA ATTGCGCGTG	2160
	GTCATGTGCT GTTTGCTCAG GCAAATTCGT TATGTGTTGC GCTAATTAAG CAACCAACAG	2220
50	TTTTAACTCA TGAGAGTAGC ATTCAATTTA TTGAAAAAGT AAAATTAAAT GATACGGTAA	2280
	GAGCAGAAGC ACGAGTTGTA AATCAA <sup>~</sup> ACTG CAAAACATTA TTACGTCGAA GTAAAGTCAT	2340
	ATGTTAAACA TACATTAGTT TTCAAAGGAA ATTTTAA <sup>~</sup> AAT GTTTTATGAT AAGCGAGGAT	2400
55		



EP 0 786 519 A2

	TTAGAAGCCG TACAAAAGGC TGTGAAGAC TTTAAAGATC TAGAAATTAT ACTTTTCGGT	2520
	GACGAAAAAA AGTATAATCT GAACCATGAA CGAATCGAAT TTAGACATTG TTCTGAAAAG	2580
5	ATTGAAATGG AAGATGAGCC TGTAGAGCG ATTAAACGTA AAAAAGATAG CTCAATGGTA	2640
	AAAATGGCTG AAGCTGTGAA ATCTGGTGAA GCAGATGGAT GTGTGTCAGC AGGTAATACT	2700
10	GGTGCTTTAA TGTCAGCTGG TTTATTCATT GTTGACGTA TTAAAGGTGT AGCTAGACCG	2760
	GCTTTAGTAG TAACATTGCC AACGATTGAT GGGAAAGGTT TTGTCTTTTT AGACGTTGGT	2820
	GCAAATGCTG ATGCTAAACC TGAACACTTA TTACAGTATG CGCAACTAGG GGATATTTAT	2880
15	GCTCAAAAAA TTAGAGGTAT TGATAATCCG AAAATCTCAT TATTAAATAT AGGAACCGAG	2940
	CCAGCTAAAG GTAATAGTTT AACGAAAAAA TCATATGAGT TATTAAATCA TGATCATTCA	3000
	TTGAATTTTG TTGGGAATAT TGAAGCGAAG ACATTAATGG ATGGCGATAC AGATGTTGTA	3060
20	GTTACCGATG GCTATACTGG GAACATGGTC CTTAAAAATT TAGAAGGTAC TGCAAAATCA	3120
	ATCGGTAAAA TGTAAAAGA TACGATTATG AGTAGTACTA AAAATAAATT AGCAGGTGCA	3180
	ATATTGAAGA AAGATTTAGC TGAATTCGCT AAAAAGATGG ATTACTCAGA ATACGGTGGT	3240
25	TCCGTATTAT TAGGATTGGA AGGTACTGTA GTTAAAGCAC ACGGTAGTTC AAATGCTAAA	3300
	GCTTTTTATT CTGCAATTAG ACAAGCGAAA ATCGCAGGAG AACAAAATAT TGTACAAACA	3360
30	ATGAAAGAGA CTGTAGGTGA AtCAAATGaG TaAAACAGCA ATTATTTTTC CGGGACAAGG	3420
	TGCCCAAAAA GTTGGTATGG CGCAAGATTT GTTTAACAAC AATGATCAAG CAACTGAAAT	3480
	TTTAACTTCA GCAGCGAACA CATTAGACTT TGATATTTTA GAGACAATGT TTAGTGATGA	3540
35	AGAAGGTAAA TTGGGTGAAA CTGAAAACAC ACAACCAGCT TTaTTGaCGC aTAGTTCGGC	3600
	ATTATTAGCA GCGCTAAAAA ATTTGAATCC TGATTTTACT ATGGGGCATA GTTTAGGTGA	3660
	ATATTCAAGT TTAGTTGCAG CTGACGTATT ATCATTTGAA GATGCAGTTA AAATTGTTAG	3720
40	AAAACGTGGT CAATTAATGG CGCAAGCATT TCCTACTGGT GTAGGAAGCA TGGCTGCAGT	3780
	ATTGGGATTA GATTTTGATA AAGTCGATGA AATTTGTAAG TCATTATCAT CTGATGACAA	3840
45	AATAATTGAA CCAGCAAACA TTAATTGCCC AGGTCAAATT GTTGTTCAG GTCACAAAGC	3900
	TTTAATTGAT GAGCTAGTAG AAAAAGGTAA ATCATTAGGT GCAAAACGTG TCATGCCTTT	3960
	AGCAGTATCT GGACCATTCC ATTCATCGCT AATGAAAGTG ATTGAAGAAG ATTTTCAAG	4020
50	TTACATTAAT CAATTGAAT GCGGTGATGC TAAGTTTCCT GTAGTTCAAA ATGTAAATGC	4080
	GCAAGGTGAA ACTGACAAAG AAGTAATTAA ATCTAATATG GTCAAGCAAT TATATTCACC	4140
55	AGTACAATTC ATTAAC TCA CAGAATGGCT AATAGACCAA GGTGTTGATC ATTTTATTGA	4200

	AACATCAATT CAAACTTTAG AAGATGTGAA AGGATGGAAT GAAAATGACT AAGAGTGCTT	4320
	TAGTAACAGG TGCATCAAGA GGAATTGGAC GTAGTATTGC GTTACAATTA GCAGAAGAAG	4380
5	GATATAATGT AGCAGTAAAC TATGCAGGCA GCAAAGAGAA AGCTGAAGcA GTAGTCGAAG	4440
	AAATCAAAGC TAAAGGTGTT GACAGTTTTG CGATTCAAGC AAATGTTGCC GATGCTGATG	4500
	AAGTTAAAGC AATGATTAAA GAAGTAGTTA GCCAATTTGG TTCTTTAGAT GTTTTAGTAA	4560
10	ATAATGCAGG TATTACTCGC GATAATTAT TAATGCGTAT GAAAGAACAA GAGTGGGATG	4620
	ATGTTATTGA CACAACTTA AAAGGTGTAT TTAAGTGTAT CCAAAAAGCA ACACCACAAA	4680
15	TGTTAAGACA ACGTAGTGGT GCTATCATCA ATTTATCAAG TGTGTGTTGA GCAGTAGGTA	4740
	ATCCGGGACA AGCAAATAT GTTGCAACAA AAGCAGGTGT TATTGGTTTA ACTAAATCTG	4800
	CGGCGCGTGA ATTAGCATCT CGTGGTATCA CTGTAAATGC AGTTGCACCT GGTTTTATTG	4860
20	TTTCTGATAT GACAGATGCT TTAAGTGATG AGCTTAAAGA ACAAATGTTG ACTCAAATTC	4920
	CGTTAGCACG TTTTGGTCAA GACACAGATA TTGCTAATAC AGTAGCGTTC TTAGCATCAG	4980
	ACAAAGCAAA ATATATTACA GGTCAAACAA TCCATGTAAA TGGTGGAAATG TACATGTAAT	5040
25	ATATTTGAGC TAAAGCTCAT TGACGCAGTG GTTGACTGGT CATCCAATGG AGAATTGTCT	5100
	GACCTAGTCA ACTTTGCGGG GGAAATTCTA AGCAACCTAG ATAAGGTTCC AGAATTTCTC	5160
	CCTAAGAAAC ACTAATCAAT aAATTGwTAA GTGTTTCTAA AATTTCTACT TGTTTTTTAG	5220
30	AATTTAAAAT GGGAAAATAT AGTAGTCTAT GTATAGGCAT TTTTAAAGGA GGTGAATCGA	5280
	CGTGGAATAT TTCGATAAAG TAAAAGATAT CATCGTTGAC CgTTTAGGTG TAGACGCTGA	5340
35	TAAAGTAACT GAAGATGCAT CTTTCAAAGA TGATTTAGGC GCTGACTCAC TTGATATCGC	5400
	TGAATTAGTA ATGGAATTAG AAGACGAGTT TGGTACTGAA ATTCCTGATG AAGAnGCTGA	5460
	AAAAATCAAC ACTGTTGGTG ATGCTGTTAA ATTTATTAAC AGTCTTGAAA AATAATAAAT	5520
40	CTTACATCTG GGTCGTCAGT ATTGTCGACT CAGTTTTTTT CTTTAATTAT CAATAGTTTT	5580
	AACGTAAAAT TAAAGATGAT TCAAGAGCAA CACATAAAGG AGATAAAATA ATGTCTAAAC	5640
	AAAAGAAAAG TGAGATAGTT AATCGTTTTA GAAAGCGCTT TGATACTAAA ATGACAGAGT	5700
45	TAGGCTTTAC TTATCAAAT ATTGATTTAT ACCAACAAGC ATTTTCGCAT TCGAGTTTTA	5760
	TTAATGATTT TAATATGAAT CGTTTAGACC ATAATGAGCG TTTAGAGTTT TTGGGTGATG	5820
50	CGGTATTAGA ATTGACGGTT TCACGATATT TATTTGATAa ACATCCCAAC TGCCAGAAG	5880
	GGAATTTAAC AAAAATGCGT GCCaCTATTG TATGTGAGCC CtCACTkGTA ATATTTGCGA	5940
	ATAAAATTGG ATTGAACGAA ATGATTTTAC TTGGTAAAGG TGAAGAGAAA ACAGGGGGAC	6000
55		

ATCAAGGACT AGATATAGTT TGGAAATTTG CTGAGAAAAGT CATTTCCTCCA CATGTAGAAC 6120  
 AAAATGAGTT ATTAGGCGTG GTAGATTTTA AAACACAATT CCAAGAATAT GTGCACCAGC 6180  
 5 AAAATAAAGG TGATGTAACC TATAATTTAA TAAAAGAAGA GGGACCGGCA CATCATCGTC 6240  
 TATTCACCTTC A 6251

10 (2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4920 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

20 ACCTACTGAA GTTGCTAATT TTTTGGAGCA ACTAAGCACT GAAATTGAAC GTCTTAAAGA 60  
 AGATAAAAAA CAACTTGAAA AAGTAATCGA AGAGAGAGAT ACTAATATTA AGTCTTATCA 120  
 AGACGTGgCA TCAATCTGTA AGTGaTGCTT TGATACAAGC TCAAAAAGCT GGTGAAGAAA 180  
 25 CTAAGCAAGC TGCAGAGAAA CAAGCTGAAG CGATTATAGC TAAGGCAGAA GCGCAAgcTA 240  
 ATcAAATGGT TGGTGACGCG GTAGAAAAAG CACGCCGTTT AGCATTCCAG ACTGAAGATA 300  
 TGAAACGTCA ATCAAAAGTA TTTAGATCGC GTTTCCGTAT GTTAGTTGAA GCGCAATTAG 360  
 ACTTATTAAA AAACGAAGAT TGGGATTACT TGTTGAATTA TGATTTAGAC GCTGAACAAG 420  
 TGACGCTTGA AAATATTCAT CATTTCATG AAAATGATT AAAGCCAGAT GAAGTTGCAG 480  
 35 CAAATGCACA AAATAATGCA TCAATACAC CAGACAATAA TCAACAATCC AATGATTCAG 540  
 AAACAACATA GAAGTAAGAA TTAAATAAAG ACAGACGCGT AATATACATT TAACTTTTCA 600  
 CAGCGAATTA GGTAATGGTG AGAGCCTAGT AAAAGCATGT ATGTTATATC ACTGGCTTTT 660  
 40 TAATATTTAA ATAATGTAAT GAGAGAACTC TAAGTTGAGT TAATAAGGGT GGTACCGCGA 720  
 GCAATCGTCC CTTTTAATTT AACTTAGAGT TTTTTAAATT TTTAAGGAGT GAAAAAATG 780  
 GATTACAAAG AAACGTTATT AATGCCTAAA ACAGATTTCC CAATGCGAGG TGGTTTACCA 840  
 45 AACAAGGAAC CGCAAATTCA AGAAAAATGG GATGCAGAAG ATCAATACCA TAAAGCGTTA 900  
 GAAAAAATA AAGGTAACGA AACATTCAAT TTACATGATG GCCCACCATA CGCGAATGGT 960  
 50 AACTTACATA TGGGACATGC CTTGAACAAA ATTTTAAAG ACTTTATTGT ACGTTATAAA 1020  
 ACTATGCAAG GGTTCTATGC ACCATACGTA CCAGGTTGGG ATACACATGG TTTACCAATT 1080  
 GAACAAGCAT TAACGAAAAA AGGTGTTGAC CGAAAGAAAA TGTCAACAGC TGAATTCCGT 1140

55

	TTAGGTGTTT GTGGTGACTT TAATGATCCA TATATTACAT TAAAACCTGA ATACGAAGCT	1260
	GCACAAATTC GTATTTTTTG AGAAATGGCA GATAAAGGTT TAATTTATAA AGGTAAAAAG	1320
5	CCAGTTTATT GGTCTCCTTC AAGTGAGTCT TCATTAGCAG AAGCAGAAAT TGAATATCAC	1380
	GATAAACGTT CAGCATCAAT TTACGTTGCA TTTGACGTTA AAGATGACAA AGGTGTCGTT	1440
10	GATGCAGATG CTAAATTTAT TATCTGGACA ACAACGCCAT GGACAATTCC ATCAAATGTT	1500
	GCGATTACCG TTCATCCTGA ATTAAAATAT GGTCAATACA ATGTAAATGG cGAAAAATAT	1560
	ATTATTGCAG AAGCCTTGTC TGACGCTGTA GCAGAAGCAC TGGaTTGGGA TAAAGCATCA	1620
15	ATCAAATTAG AAAAAGAATA CACAGGTAAA GAATTAGAGT ATGTTGTAGC ACAACATCCA	1680
	TTCTTAGACA GAGAATCGTT AGTGATTAAT GGTGATCATG TTACTACAGA TGCTGGTACA	1740
	GgTTGTGTAC ATACAGCACC AGGTCACGGG GAAGATGACT ATATTGTTGG TCAAAAATAT	1800
20	GAATTGCCAG TAATTAGTCC AATCGATGAT AAAGGTGTAT TTACTGAAGA AGGCGGCCAA	1860
	TTTGAAGGGA TGTTCATGA TAAAGCTAAT AAAGCCGTTA CTGATTTATT AACAGAAAAA	1920
	GGTGCACTAT TAAAATTAGA CTTTATTACA CATAGCTATC CACACGACTG GAGAACAAAA	1980
25	AAACCTGTAA TCTTCCGTGC TACACCACAA TGGTTTGCCT CAATCAGTAA AGTAAGACAA	2040
	GATATTTTAG ATGCAATCGA AAATACAAAC TTCAAAGTAA ATTGGGGTAA AACACGTATT	2100
30	TACAATATGG TTCGTGACCG TGGCGAATGG GTTATTTCTC GTCAACGTGT GTGGGGTGTA	2160
	CCGTTACCAG TATTTTATGC TGAAAATGGC GAAATTATCA TGACGAAAGA AACAGTGAAT	2220
	CATGTTGCTG ATTTATTTGC AGAACACGGT TCAAATATTT GGTTTGAAAG AGAAGCGAAA	2280
35	GACTTACTAC CAGAAGGATT TACACATCCA GGCAGCCCTA ACGGTACATT TACTAAAGAA	2340
	ACAGACATTA TGGACGTTTG GTTTGATTCT GGTTCATCAC ACCGTGGCGT GTTGGAAACA	2400
	AGACCGGAAT TAAGTTTCCC AGCGGATATG TATTTAGAAG GTAGTGACCA ATATCGTGGT	2460
40	TGGTTCAACT CTTCTATCAC AACTTCAGTT GCTACAAGAG GAGTATCACC TTATAAATTC	2520
	TTACTTTCTC ATGGTTTTGT TATGGACGGT GAAGGTAAGA AAATGAGTAA ATCTTTAGGT	2580
45	AATGTGATTG TACCTGACCA AGTGGTTAAA CAAAAGGTG CTGATATTGC GAGACTTTGG	2640
	GTAAGTAGTA CGGACTATTT AGCTGATGTT AGAATTTCTG ATGAAATTTT AAAACAAACA	2700
	TCTGATGTTT ATCGTAAAAT CAGAAATACA TTAAGATTTA TGTTAGGTAA CATTACGAT	2760
50	TTCAATCCTG ACACAGATAG CATTCCTGAA TCAGAGTTAT TAGAAGTGGA TCGTTACTTG	2820
	CTAAATCGTT TACGTGAATT TACTGCAAGT ACGATTAAACA ACTATGAAAA CTTTGACTAC	2880
	TTAAATATTT ATCAAGAAGT TCAAACTTT ATCAATGTTG AGTTAAGTAA TTTCTATTTG	2940
55		

	CAAACAGTGT TATATCAAAT TTTAGTTGAT ATGACGAAGT TGTTAGCACC AATCTTAGTG	3060
5	CATACAGCTG AAGAAGTTTG GTCTCATACA CCACATGTTA AAGAAGAAAG TGTTCACTTA	3120
	GCAGACATGC CTAAAGTTGT AGAAGTAGAT CAAGCTTTAT TGGATAAATG GCGTACATT	3180
	ATGAATTTAC GTGATGATGT GAACCGTGCA TTAGAACTG CTCGTAATGA AAAAGTTATT	3240
10	GGTAAATCAT TAGAAGCTAA AGTTACGATT GCTAGTAACG ATAAATTTAA TGCATCTGAA	3300
	TTCTTAACTT CATTTGATGC ATTACATCAA TTATTTATCG TGTCACAAGT TAAAGTTGTA	3360
	GATAAGTTAG ACGATCAGGC AACAGCTTAT GAACATGGTG ATATTGTCAT CGAACATGCA	3420
15	GATGGTGAAA AATGTGAAAG ATGTTGGAAC TATTCAGAGG ATCTTGGTGC TGTTGATGAA	3480
	TTGACGCATC TATGTCCACG ATGCCAACAA GTTGTAATAAT CACTTGTATA ATTGAAATTG	3540
	TATAAAGTAC TCATACAGAT GATATAAATT AAAGCTCTCT TCATAATCAT GTTGTAGTTT	3600
20	TTGTTGACAT GATGAAGAGA GTTTTTTTGT GAATAAAAAA ATGACCAAGT TACCGGTCAT	3660
	ATATGTAAAA AATGTGCGAT TTAATAAAT AAAAATTATT CAGGAATGGT ACAAATTCTC	3720
25	TGAGGCATAT AAATGCGTTA TAGTTGCTAT TCTCAATTAT GTTCGCGATA ATTTTAAGTA	3780
	AAAGTAAGCA CAGATATTGA ATTTGATAGG AGTTAATTGA ATGTATCATA ACAGTAACGC	3840
	AAACTTTGTC AATGGTATCA CTTTAAATGT GAGAGATAAG AATGAATTAA AGCCATTTTA	3900
30	TGAGGACATA TTAGGATTAA ATATTATAAA TGAGACATTA ACATCGATAC AATATGAAGT	3960
	AGGTCAAAAT AATCATGTCA TTACACTTGT TGAATTACAA AATGGACGTG AACCTTTAAT	4020
	GTCCGAAGCG GGAAGTTTC ATATCGCAAT TAACTACCT CAAATTAGTG ATTTAGCTAA	4080
35	TTTACTAATT CATTTAAGCG AATATGATAT TCCAGTTAAC GGAGGTATAC AGCCTGCTTC	4140
	GTTATCATTA TTTTTTGAAG ACCCGGAAGG AAACGGTTTT AAATTTTATG TTGATAAAGA	4200
40	CGAAGCGCAA TGGACGAGGC AAAATAATTT AGTAAAAATT GATATTAGAC CATTAAATGT	4260
	ACCGAGATTA GTGAGTCATG CAACAAAATT GTTATGGTTA GGTATTCCAG ATGACGCTAT	4320
	TATAGGTGCA TTGCATATTA AGACAATTCA TTTATCAGAG GTAAAAGAGT ACTACCTCGA	4380
45	TTATTTTGA TTAGAGCAAT CGGCATATAT GGATGATTAT TCAATATTTT TAGCATCGAA	4440
	TGGCTATTAT CAACATTTGG CCATGAATGA TTGGGTATCA GCAACGAAAC GTGTAGAAAA	4500
	TTTTGATACG TATGGATTAG CAATTGTTGA CTTTCATTAT CCTGAAACAA CACATTTAAA	4560
50	TTTACAAGGT CCGGATGGTA TCTATTATCG CTTTAATCAT ATCGAAGTTG AAGATTAGTA	4620
	TATACTTTGA ATGGACGAAC CATATAATGA ATCGTTTTTA ATGATCTTTT TATACAAGTT	4680
55	ATGAAGGAGG CTGGGACATT AAGTTCCTAG GCAATGTAAA AAGCTGATTT CTATTAATTA	4740

TTTTCTTAT ATTAATTGCC ATTAATACAA AACCTAGCTC TCGTTTAACT TTATTTATTC 4860  
 CTCGAACTGA CATTCCGnGTG AACTCAAAAT nGCCTACTTn CTTAAATTAC CAATATCTAT 4920

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 626 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TGGATTGCCA TTACATGGAC AAGATTTAAC TGAATCAATT ACACCATATG AAGGTGGTAT 60  
 CGCTTTTGCA AGTAAACCAT TAATTGATGC TGATTTTATT GGTAAATCTG TATTAAAAGA 120  
 TCAAAAAGAA AATGGTGCAC CAAGAAGAAC AGTGGGATTA GAATTACTTG AAAAAGGAAT 180  
 TGCAAGAACT GGTTATGAAG TTATGGATTT AGATGGAAAT ATTATTGGAG AAGTAAC TTC 240  
 AGGAACACAG TCTCCATCAT CAGGAAAATC AATTGCACTT GCAATGATAA AAAGAGATGA 300  
 GTTTGAAATG GGTAGAGAGT TGCTTGTTC AAGTTCGTAAG CGTCAATTAA AAGCGAAAAT 360  
 TGTTAAGAAA AATCAAATTG ATAAATAATT AAAAAGGGGT GTGCATTGTG AGTCATCGTT 420  
 ATATACCTTT AACTGAAAAA GACAAGCAAG AAATGTTACA AACCAATTGGT GCAAAATCTA 480  
 TAGGAGAATT ATTCGGTGAT GTACCAAGTG ACATTTTATT AAATAGAGAT TTAAATATTG 540  
 CTGAAGGCGA ACGGAGAACA ACGTTACTTA GAAGATTnAA TCGCATTGCA AGCAAGAGTA 600  
 TCACTAGAGG AACGCGTACA TCGTTT 626

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1126 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

nGGAAGTGGT GTATATATTT GTAATGAGTG TATTGAATTA TGCTCAGAAA TCGTCGAAGA 60  
 AGAATTAGCT CAAAACACTT CTGAAGCGAT GACAGAATTA CCTACTCCTA AAGAAATTAT 120  
 GGATCATTTA AACGAATATG TTATTGGTCA AGAAAAAGCT AAAAAATCTT TAGCTGTAGC 180  
 TGTTTATAAC CACTATAAGC GTATTCAACA ATTAGGACCA AAAGAAGATG ATGTTGAATT 240

AACCTTAGCC AAGACGTTGA ATGTACCATT TGCAATTGCA GATGCGACAA GTTTAACTGA 360  
 AGCTGGTTAT GTAGGCGATG ATGTTGAAAA TATCTTGTG AGATTAATTC AAGCAGCTGA 420  
 5 CTTTGACATT GATAAAGCCG AAAAAGGTAT TATTTATGTA GATGAAATTG ATAAAATTGC 480  
 ACGTAAATCT GAAAACACAT CTATAACACG TGACGTTTCA GGTGAAGGTG TTCAACAAGC 540  
 10 ATTGCTTAAA ATCTTAGAAG GTACGACTGC AAGTGTTCCG CCACAAGGTG GACGCAAACA 600  
 TCCAAACCAA GAAATGATTC AAATTGATAC AACAAATATC TTATTTATTC TTGGTGGTGC 660  
 CTTTGATGGT ATTGAAGAAG TGATTAAGCG CCGTCTTGGT GAAAAAGTTA TTGGTTTCTC 720  
 15 AAGCAATGAA GCTGATAAAT ATGACGAACA AGCATTATTA GCACAAATTC GCCCAGAAGA 780  
 TTTGCAAGCC TATGGTTTGA TTCCTGAATT TATCGGACGT GTGCCAATTG TAGCTAATTT 840  
 AGAAACATTA GATGTAAC TGTTGAAAA CATCTTAACG CAACCTAAAA ATGCACTTGT 900  
 20 GAAACAATAT ACTAAAATGC TGAATTAGA TGATGTGGAT TTAGAGTTCA CTGAAGAAGC 960  
 TTTATCAGCA ATTAGTGAAG AAGCAATTGA AAGAAAAACA GGTGCGCGTG GTTTACGTTT 1020  
 AATCATAGAA GAATCGTTAA TCGATATTAT GTTTGATGTG CCTTCTAACG AAAATGTAAC 1080  
 25 GAAGGTAGTT ATTACAGCAC AAACmATTAA TGrAGaACTG AACCAG 1126

## (2) INFORMATION FOR SEQ ID NO: 29:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4392 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

ATTGACTTCT TAGCAATnAA TaTGAGTGAA GAACGTACTG TTGAAGTACC AGTTCAATTA 60  
 40 GTTGGTGAAG CAGTAGGCGC TAAAGAAGGC GGCGTAGTTG AACCAACCATT ATTCAACTTA 120  
 GAAGTAACTG CTA CTCCAGA CAATATTCCA GAAGCAATCG AAGTAGACAT TACTGAATTA 180  
 45 AACATTAACG ACAGCTTAAC TGTGCTGAT GTTAAAGTAA CTGGCGACTT CAAAATCGAA 240  
 AACGATT CAG CTGAATCAGT AGTAACAGTA GTTGCTCCAA CTGAAGAACC AACTGAAGAA 300  
 GAAATCGAAG CTATGGAAGG CGAACAACAA ACTGAAGAAC CAGAAGTTGT TGGCGAAAGC 360  
 50 AAAGAAGACG AAGAAAAAAC TGAAGAGTAA TTTTAATCTG TTACATTAAA GTTTTTTATAC 420  
 TTTGTTTAAC AAGCACTGTG CTTATTTTAA TATAAGCATG GTGCTTTTGT TGTATTATA 480  
 AAGCTTAATT AAACCTTATT ACTTTGTACT AAAGTTTAAT TAATTTTAGT GAGTAAAGA 540  
 55

	CTTACTAAGC TAAAGAATAA TGATAATTGA TGGCAATGGC GGAAAATGGA TGTGTGTCATT	660
	ATAATAATAA ATGAAACAAT TATGTTGGAG GTAAACACGC ATGAAATGTA TTGTAGGTCT	720
5	AGGTAATATA GGTAAACGTT TTGAACTTAC AAGACATAAT ATCGGCTTTG AAGTCGTTGA	780
	TTATATTTTA GAGAAAAATA ATTTTTCATT AGATAAACAA AAGTTTAAAG GTGCATATAC	840
10	AATTGAACGA ATGAACGGCG ATAAAGTGTT ATTTATCGAA CCAATGACAA TGATGAATTT	900
	GTCAGGTGAA GCaGTTGCAC CGATTATGGA TTATTACAAT GTTAATCCAG AAGATTTAAT	960
	TGTCTTATAT GATGATTTAG ATTTAGAACA AGGACAAGTT CGCTTAAGAC AAAAAGGAAG	1020
15	TGCGGGCGGT CACAATGGTA TGAAATCAAT TATTAAATG CTTGGTACAG ACCAATTTAA	1080
	ACGTATTCGT ATTGGTGTGG GAAGACCAAC GAATGGTATG ACGGTACCTG ATTATGTTTT	1140
	ACAACGCTTT TCAAATGATG AAATGGTAAC GATGGAAAAA GTTATCGAAC ACGCAGCAGC	1200
20	CGCAATTGAA AAGTTTGTG AAACATCAGC ATTTGACCAT GTTATGAATG AATTTAATGG	1260
	TGAAGTGAAG TAATGACAAT ATTGACAACG CTTATAAAG AAGATAATCA TTTCAAGAC	1320
25	CTTAATCAGG TATTTGGACA AGCAACACA CTAGTAACTG GTCTTTCCCC GTCAGCTAAA	1380
	GTGACGATGA TTGCTGAAAA ATATGCACAA AGTAATCAAC AGTTATTATT AATTACCAAT	1440
	AATTTATACC AAGCAGATAA ATTAGAAACA GATTTACTTC AATTTATAGA TGCTGAAGAA	1500
30	TTGTATAAGT ATCCTGTGCA AGATATTATG ACCGAAGAGT TTCAACACA AAGCCCTCAA	1560
	CTGATGAGTG AACGTATTAG AACTTTAACT GCGTTAGCTC AAGGTAAGAA AGGGTTATTT	1620
	ATCGTTCCTT TAAATGGTTT GAAAAAGTGG TTAACCTCTG TTGAAATGTG GCAAAATCAC	1680
35	CAAATGACAT TCGGTGTTGG TGAGGATATC GATGTGGACC AATTCTTAA CAAATTAGTT	1740
	AATATGGGGT ACAAACGGGA ATCCGTGGTA TCGCATATTG GTGAATTCTC ATTGCGAGGA	1800
	GGTATTATCG ATATCTTCC GCTAATTGGG GAACCAATCA GAATTGAGCT ATTTGATACC	1860
40	GAAATTGATT CTATTCGGA TTTTGATGTT GAAACGCAGC GTTCCAAAGA TAATGTTGAA	1920
	GAAGTCGATA TCACAACTGC AAGTGATTAT ATCATTACTG AAGAAGTGAT CAGCCATCTT	1980
45	AAAGAAGAGT TAAAACTGC ATATGAAAAT ACAAGACCCA AAATAGATAA ATCAGTGGCG	2040
	AATGATTTGA AAGAAACGTA TGAAAGCTTT AAATTATTCTG AAAGTACATA CTTTGATCAT	2100
	CAAATACTAC GTCGCTTAGT AGCGTTTATG TATGAAACAC CTTGACAAT TATTGAGTAT	2160
50	TTCCAAAAAG ATGCAATCAT TGCAGTTGAT GAATTTAATC GTATTAAAGA AACTGAAGAA	2220
	AGTTTAACAG TAGAGTCTGA TTCGTTTATT AGCAATATTA TTGAAAGTGG TAATGGATTT	2280
55	ATAGGACAAA GTTTTATAAA ATATGATGAT TTTGAAACAT TGATTGAAGG CTATCCTGTC	2340



	TCATGTAAAC CTGTCCAACA ATTTTATGGG CAATATGACA TTATGCGTTC TGAATTTCAA	2460
	CGATATGTTA ATCAAACTA TCATATCGTG GTTTTGGTCG AAACCGAAAC TAAAGTTGAA	2520
5	CGTATGCAAG CGATGTTAAG TGAAAtGCAT ATTCCATCAA TAACAAAATT GCATCGCTCA	2580
	ATGTCATCGG GGCAAGCAGT GATTATTGAA GGCAGTTTAT CTGAAGGATT TGAACCTACCT	2640
10	GATATGGGAT TAGTTGTCAT TACTGAGCGT GAgcTTTTTA AATCAAAACA GAAAAAGCAA	2700
	CGAAAACGTA CGAAAGCTAT CTCAAATGCT GAAAAAATTA AGTCTTACCA AGATTTTAAAT	2760
	GTGGGAGATT ATATTGTTCA TGTGCATCAT GGTGTTGGTA GATATTTAGG TGTGAGACG	2820
15	CTCGAAGTGG GGCAAAACGCA TCGTGATTAT ATTAAATTGC AATATAAAGG TACGGATCAA	2880
	CTATTGTGTC CAGTAGATCA AATGGATCAA GTTCAAAAAT ATGTAGCTTC GGAAGATAAG	2940
	ACGCCAAAAT TAAATAAACT CGGTGGCAGT GAATGGAAAA AAACAAAAGC TAAAGTTCAA	3000
20	CAAAGTGTG AAGATATTGC TGAAGAGTTG ATTGATTTAT ATAAAGAAAG AGAAATGGCA	3060
	GAAGGTTATC AATATGGGGA AGACACAGCT GAGCAAACAA CATTGAATT AGATTTTCCA	3120
	TATGAACTTA CGCCTGACCA AGCTAAATCT ATCGATGAAA TTAAAGATGA CATGCAAAAA	3180
25	TCGCGTCCAA TGGATCGCTT GCTATGTGGT GATGTTGGTT ATGGTAAAC TGAAGTTGCA	3240
	GTGAGAGCAG CATTCAAAGC TGTAATGGAA GGAAAGCAGG TTGCATTTT AGTTCCTACA	3300
30	ACTATTTTAG CTCAGCAACA TTATGAGACG TTAATTGAGC GTATGCAAGA TTTTCCTGTT	3360
	GAAATTCAAT TAATGAGTCG TTTTAGAACG CCTAAAGAGA TAAACAAAC TAAGGAAGGA	3420
	CTTAAACTG GATTTGTGA CATAGTTGTT GGTACACACA AATTACTTAG TAAAGATATA	3480
35	CAGTATAAAG ATTTAGGGCT GTTGATTGTA GATGAAGAAC AACGATTTGG TGTACGCCAT	3540
	AAAGAGCGTA TTAAAACATT AAAACATAAT GTAGATGTAC TAACATTGAC TGCAACCCCA	3600
	ATAGCTAGAA CATTGCATAT GAGTATGCTA GGTGTGCGGG ATTTGTCAGT GATTGAAACG	3660
40	CCGCCAGAAA ATCGTTTCCC AGTTCAAACA TATGTATTAG AACAGAACAT GAGTTTTATC	3720
	AAAGAAGCTT TAGAAAGAGA ACTATCCCGT GATGGCCAAG TGTTTTATCT TTATAATAAA	3780
45	GTGCAATCCA TTTATGAAAA ACGAGAACAA CTCCAGATGT TAATGCCAGA TGCTAACATT	3840
	GCAATTGCTC ATGGACAAAT GACAGAGCGC GATTTAGAAG AAACGATGTT AAGTTTTATC	3900
	AATAATgAAT ATGATATTTT AGTAACGACG ACGATTATTG AAACAGGTGT CGATGTCCCA	3960
50	AATGCAAATA CTTTGATCAT TGAAGATGCA GATCGCTTTG GATTGAGTCA GTTGATCAA	4020
	TTAAGAGGTC GTGTTGGTCG TTCAAGTCGT ATTGGTTATG CATACTTCTT ACATCCAGCA	4080
	AATAAGGTAC TAACTGAGAC TGCAGAAGAT CGATTACAAG CGATTAAAGA ATTTACGGAG	4140
55		

TTAGGTAAAC AACAGCACGG CTTTATTGAT ACAGTTGGAT TTGATTTGTA CAGTCAAATG 4260  
 TTAGAAGAAG CTGTAAATGA AAAACGTGGT ATTAAGGAAC CAGAATCTGA GGTGCCAGAA 4320  
 5 GTCGAAGTTG ATTTAACTT GGATGCATAT TTGCCAACAG AATATATTGC AAATGAACAA 4380  
 GCTAAAATTG AA 4392

(2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 729 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

20 TTCTTTTGA ATCTATATCG AGGTGGTTGG TAGGTTTCATC TAAAATAAGT ACATTGTCAC 60  
 GTTGCAACAT AAGTAGTGCT AGTTGTAAAC GTGCTTTTTTC ACCACCAGAT AAATCATTAA 120  
 TTATCTTTTT AACATCGTCT TGTACAAATA AGAAACGTCC AAGAACTGCT CGAATATCTT 180  
 25 TTTCATTCAT TAACGGATAT TGATCCACA CATAATCTAA AATCGTTTTA CTAGATTTAA 240  
 ATTCTGCTTG CTTTGTATCA TAATAACCAA TTTGTAAATT TGCGCCGAAA GTAATATCGC 300  
 CATTAAGCGC TTTTGTGTA TTAGCAATAG TTTTAATTAA GGTCGATTTT CCAATACCAT 360  
 TTGGCCCAAT GATTGCTATA TGATCGCCTT TAGAGACCTC TATACTCATA GGTTTGGTAA 420  
 TTGCAGTTTG ATAACCGATT TCTAAATTTT TTACATGCAT GACGTCATTA CCTGTATTCC 480  
 35 GGTCAAAGCC AAATTGAATA TTGCACTTT TGGCATCTAA CATTGGTTTA TCAATGCGTT 540  
 CCATTTTTTC TAAAATCTTA CGTCTACTTT TTGCCATTCC ACTTGTGAA GCACGGGTAA 600  
 TATTTTCTC AACAAAAGTT TCTAATCGTT TTATTTCTGC TTGTTGACTT TCATATTCTT 660  
 40 GCATTCGTTT TTGATAATAT AAATCCCGTT GCTGTATAAA TTCCTCGTAA TTACCAACAT 720  
 AGCGTTTGA 729

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13856 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

	TGATGTTTTCG ATACATTTGT TGCACCTTGT GGATATACTT TAAAGGTTGT GTCGTATGTT	120
	TCCTTACTAT CTTTAGCTTC AGATTCTGT GATTCAACCG TTTTATATTT TTCAAGTGCA	180
5	TGTCCTTCAA TATCAACTCG TGAATAATG CGATTCAACC ATGCTGGTAA ATACCACGAA	240
	CCTTTtCCAA ACAATTTcGt TAATGCAGGA ATTAACATCA TtCTGACTAC GAAGGCATCA	300
10	AAGAGTACAC CAAACGCTAA TGCCATACCC ATTGATTTAA TCATGACATC TTCTTGGAAT	360
	ACAAACGCAA AGAAGACACT AAACATAATT AATGCAGCTG CTACAATAAC AGGACCGCTT	420
	TCTTTCAATC CTACTTTGAT AGAATAATCA TTATCCCCTG TTTTACTATm yyCTTCATGr	480
15	ATTCGCGACA TAAGGAAGAC TTCATAATCC ATCGCTAATC CAAATAAGAT ACCTATAGTA	540
	ATAACCGTA AAAATGCTAG CATTGGTCCT GTCGTTTCAA TACCAAACAG ACCTTTCATA	600
	AAACCATCTT GCATTACTAA TGTTGTAAAT CCTAATGTTG CCATTAAATGA CAAGACGAAT	660
20	CCTAAACTG CTTTTAATGG TATTAGAATT GAACGGAAGA CAATCATTAA TAAGAAAAAT	720
	GCTAATACAA CAATGACTGA GGCAAATAAA GGTATCGCCT CATTTAACTT TTTAGACATA	780
	TCAATATTAA TGACACTTTG TCCCGAAATC TCCGTTTTGA ACCCATATTT ATCTTGTGCA	840
25	TCTTTATGAT AATCTCGTAA ATCATGCACT AAATCATTG TACTCTCTGC ATTAGGCCCT	900
	TGCTTAGGTA TCACGACCAT CAAAGCGTAA TCATTATCTT TACTCATTG TGGTGGCGTA	960
30	ACGATATCTA CATTTTTCTT ATCTTTAATA TCTTTATATA CAGACTGTAA ATCTTGTTGT	1020
	AATCCTTGTG GATCATCCTT TTTATCTTTC ACATTTATCA ACATCGGTAT TTGGCCATTA	1080
	AATCCTTCAC CAAATTTATC CGAGATAATA TCGTAAGCTT TTTTCTGTGT AGAATCTGCT	1140
35	GGTTTAAACAC CGTCATCTGG AATACCAAGT CGCATATGAC TAACTGGTAT TGCAGCTGCT	1200
	ACTAATATGA TTAAACCTAG TAATACTGCC GCAAGTGCAT TTCCTGTAAT AAATTTAGAC	1260
	CATGGCGTAT CAATATCTTT TTTGAATTTA GACTGTAATT TATTCATTT AATGCGTTtA	1320
40	TGGAATAATGC TTATTAATGC AGGTAATAAA GTTAAAGCGC TAAGTACTGC AAAACAACA	1380
	CTAATTGCCG AAGCAAATCC CATTACCGCT AAGAAGTCAA TGCCTACTAA TGATAAACCA	1440
45	CATACTGCAA TTACAACGT TACACCAGCA AAAACAACCTG CACTACCTGC TGTTCTATT	1500
	GCAAGACCAA TGCCTTTAAT GTAATCTGTT TCAGTTTTCA TAACTTGTCG ATATCTGAAT	1560
	AAAATAAATA ATGCATAATC GATACCAACT GCTAGTCCAA TCATTACGGC TAATGTCAGT	1620
50	GTGACATTTG GTATATCGAA TGCATAAGTT AACAACTGA TAATACCTAC ACCAGAGGCT	1680
	AGACCAATCA ATGCACCTAT AATTGGTAAT CCTGCAGCAA TGAAGTGAACC GAATGTGATT	1740
55	AACAGTACAA CAAATGCAAC AATAATACCA ACTAGTTCAG AATTACCGCC TACTTCTGTA	1800

	AAATGACTTT TAACATTATC TCTAGAGCCA TCTTTTAAAG ATGTTTGAAT AACGTCATAT	1920
	GTGATATCTG CAAATGCAGT TGTTTTATCT TTTACTAATTT GCTTATTTTC ATAAGGATCT	1980
5	GATATTTTAT CAATGTGCTT GTCATCTTTT TTAATATCAT CTAACGTTTT CTTAATATCT	2040
	TTAGTAATGT TCGGTTGCAC AATACCATCA TCTTTAGTCG TCTTAAAGAC AACACGTATT	2100
10	TGTGCCTTTT CACTATCTTG ATTAAAATGT TTTTCAATCT TTTTATTCGT ATCTAACGAC	2160
	TCTAATCCTG TCATTTTAAT ATCATTGTCA AATTTCCGGT CATTGTAGC AAGTGGTATC	2220
	AATATTGCAG CTACAATCAC TATCCATGCA ATGACCGCGG ACCATTTATG TTTTGCGATG	2280
15	AATGTCCCCA TCTTATATAA AAATTTTGCC AAAGTATATT GCCTCCTTTT AAAATCAACG	2340
	TTATAGTTTA AATATACAGT GTAGATTATT GTTCGATTAT AGTATCTATC CCCGACCTCT	2400
	TAAAGAATCA ATTGGAAAAT TTTGTATATT AAACACACA CAAAGGAGAA ATGTAGATGA	2460
20	AAGAGACTGA TTTACGAGTT ATAAAGACAA AAAAAGCATT GTCGAGTAGC TTGCTACAAT	2520
	TGTTAGAACA GCAATTATTC CAAACGATTA CTGTCAATCA AATTTGCGAC AACGCACTCG	2580
25	TACACCGTAC AACATTTTAT AAACATTTTT ATGATAAATA TGATCTTCTA GAGTACTTGT	2640
	TCAATCAATT GACTAAAGAC TACTTTGCTA GAGATATCAG TGACCGTCTT AATCATCCAT	2700
	TCCAAACGAT GAGTGATACG ATTAATAATA AAGAGGATTT GAGAGAAATC GCAGAATTCC	2760
30	AAGAAGAAGA CGCTGAATTT AATAAAGTAT TAAAAAATGT CTGCATTAAA ATTATGCATA	2820
	ACGATATCAA AAATAATAGA GACCGTATCG ATATTGACAG CGACATCCCA GATAATCTCA	2880
	TATTTTATAT TTATGACTCG TTGATTGAAG GTTTTATACA TTGGATAAAA GATGAAAAAA	2940
35	TTGATTGGCC TGGCGAAGAT ATTGATAACA TTTTCCATAG ATTAATCAAT ATTAAGATTA	3000
	AATAGTAGAT GAGAACTCA TGAGCGTTAC CAACATTCAT AATAAAAACG ATAGTGKACA	3060
	CGTTAATGAA TTCGTGTACT ACTATCGTTT TTTATTTTTA TCGTGCTTAT CGCTATTAAA	3120
40	ACAACGTATA CACAACACAT AAACATGAA GAAAAAATA AATCCGCTAT CTAAATGACT	3180
	TTGACTCAGT TGTTTAAATG ACCAAATTGC TAATACAATT CCCATTATTA TTGAAATAAC	3240
45	GTATCTCACA TTCTTATACC TATAATCCTT TTCTAAAAAT ATGGTTGCTA TTAATTAATT	3300
	TTTAAAGTTA TAAATAAAAA GAGCCAACCG CAATGGATGG CCCTTGTTCA TTATGAAGCA	3360
	TTAGAACATT TCTGAAACAA CCTTTTGTTT TAAGAAGTGT AATAAGTAGT CTGGACTACC	3420
50	TGTTTTAGCG TCCGTACCTG ACATTTTGAA ACCACCAAAT GGATGGTATC CAACAACGTC	3480
	TGAAGTACAG CCTCTGTTAA GGTATAAATT GCCTACATCA AATTCGTTTA CCGCTTTAAT	3540
55	CCAATGCTCG CGATTATTTG TAATCACTGC ACCAGTTAAA CCGTAATCTG TATCATTTGC	3600

	TTCTTCTTGC ATGATTCTAT CTTTAGATTT AAGTCCTGAA ATGATTGTTG GTTCTACAAA	3720
	GTAACCTTTT GAATCATCAG TGCCGCCACC TTGTTCTAAT TTACCTTCTT CTTTACCAAT	3780
5	CTCAATATAA TTTTAAATCT TATCAAATTG TTTTATTATTA ATAAGTGGGC CCATATACGT	3840
	ATTGTCTACA GTATTGCCCA ACGTTAATTC TTTTGTTAAT TTGATTGATT TCTCTAATAC	3900
10	TTCGTCATAA ACGTCTTTAT GCACAATTGC ACGTGAACAT GCTGAACATT TTTGACCAGA	3960
	AAAACCAAAT GCTGACGTTA CAATAGCTTC TGCTGCCATA TCTGTATCAA TATTTTCATC	4020
	AACTACAATG GCATCTTTAC CACCCATTTT AGCGATAACA CGTTTCAAGA AGTTTTGACC	4080
15	TTCTTGAACA ACGGCACTAC GTTCATAAAT TCTAGTACCT GTCGCACGTG ATCCTGTAAA	4140
	TGTAACGAAA TCGGTATCTT TATGATCAAC TAAGTAATCA CCAATTTCTT TCGGATCACC	4200
	AGGAACAAAG TTAAGTACGC CTTTGGTAA TCCTGCTTCT TCTAAAATTT CCATTAATTT	4260
20	ATAAGCGATA TAAGGTGTAT CCTCAGCAGG TTTCAATAAC ACTGTATTAC CTGCCACAAC	4320
	TGGTGCTAAA GTTGTACCAG CCATAATCGC AAACGGGAAG TTCCACGGCG GAATTGTAAC	4380
	ACCTGTACCA ATTGATTTAT AGAAATATTT ATTGTGTTCA CCTTCACGAT CAAGTACTGG	4440
25	CTTACCTTGA GCCAAGTCCA TCAATGAACG TGCATAGTAT TCAATAAAAT CAATACCTTC	4500
	AGCTGCATCA CCAACTGCTT CATCCCATGG CTTACCTGCT TCATAAACCA TAATTGCTGC	4560
30	AATTTCCGCT TTTCGACGAC GAATAATTGC CGAAACACGT AACATAAGCT CTGCACGATC	4620
	ATTTGCTGAC CATGTTTTCC AAGATTTATA AGCTTCGTTT GCTGCTTTAA ACGCATCTTC	4680
	AACATCTTGT TTTGTTGCCT TTGATGCATT TGCAATCACT TGTGATGTGT CTGCAGGATT	4740
35	GATTGATTTA ATTTTGTGAT CTTTGAAAAT CTTCTCTCCA TTAATCACTA ATGGTATGTC	4800
	TTGACCTAAT TCTTTTTCCA CGTCTTTCAA TGCTTTCTTA AACATATCCA CATTTTCTTG	4860
	GACTGAAAAA TCGTAACCAG GTTCATTTTT AAATTCTACT ACCATGTACA CTTACCCCTT	4920
40	ATAAATTTTG AAAGTGGTTT AACCCTTTGA TTTAATGATA TAACATCATT TAAACTCATT	4980
	TTACTATGAT TAAGGTTAGT TTTGCAATCG CTTTCATTTT TATGTTTTAT CACTTATTCT	5040
45	CAAGTATTTT GAAATTGATT GGTTACTTTT TAAATTTAT ATGGGTCGCA ACTGCTACTT	5100
	TATCGTTTCG TCATTTAATG TTTCGGATGG TAGGTCATTA TCAATTTTAC GAACGACTTT	5160
	ACAAGGGTTT CCAACCGCTA AGCTGTGTGG CGGAATATCT TTAGTGACAA CACTACCAGC	5220
50	ACCAATCACA CTGCCTTCTC CAATCGTCAC CCCTGGTAAC ACGGCTACAT GACCGCCAAA	5280
	CCAAGTATTA CTGCCAATAT GAATGGGTCC GGCTTTTTC A AACCTTCAT TTCTATGATG	5340
	GAAATTAAGT GGATGTGTCG CTGTGTAGAA TCCACAATTA GGTCCTATAA AAACATTATC	5400
55		

	TCCTAGTTTA	ACGTTCCAAC	CATAATCTGT	ATCAAAAGGA	ATCGAAATAC	TTACATTGTC	5520
	TGTTGTTGTT	TGAAATAATT	GATCAATTAA	TTCTTTTCTT	TTATTTGTAG	CACTCGGTCT	5580
5	TGTATGATTT	AATTCAAAGC	AAATATCTTT	CGCTCGTGCA	CGTTCATTGA	TTAAGTATTG	5640
	ATCAAAGTTT	GCATCGTACC	ATTTTCTGTC	TAACATTTTT	TCTTTTTCAG	TCATTACACC	5700
10	TTTCAACTCC	TAATAACTTA	TTTACTTGTT	TAAAAGTTAA	TCAAATAAAC	CTTCGCCTAT	5760
	GCAACTAATA	CGCTATAACA	TTATGAAATC	ATGACCTTAT	CACCCTTATC	TATACAATTC	5820
	TCGCATCAAA	TACTGCTAAA	GTAGTAGATA	AATTCAATAC	TACAGACGCA	TTCATTTTTT	5880
15	AATCTATTAA	CGTACAATGT	GAGTAAGAGA	AATATAAAGG	AGTATGATAG	CGATGAGAAT	5940
	ATTAATTACA	GGCACAGTTG	CTATCTTAAT	CATTCTAGGT	TTGGTCAAAA	CGATACAAGA	6000
	TTACGAAATG	ACAAACGACA	CGAGTCGTCA	GTTGTCTAGC	AACAAAGATG	ATGATAAAGT	6060
20	CATCCATCTT	AATAATTTTA	AAAATTTACA	TGCGAAAGAA	TTTAACCCAT	CTGATTTCTT	6120
	TTAAGTCACC	TAAGAATTGC	AAATCCAGAA	GTCATTTAAG	TTTTACCTTT	CATTTCATACA	6180
25	TCCTTTAATA	TTAATTACGA	CTTCTTTTAT	ATAGATGCTA	AGTAGAGAGA	TTGTTGTGCA	6240
	ATGTTTGAC	GGCAATCTCT	CTTTTTCTTT	TTAAAATTGG	TAAAAGTAAA	ACGCAACGAT	6300
	TGACTTATAT	ACCTATAGGG	GGTACATTAG	ACGTGTAACA	ATGAATCACA	GGGAGGCAAT	6360
30	AATGTGGCTA	ATACGAAAAA	AACAACATTA	GATATCACTG	GTATGACTTG	TGCCGCATGT	6420
	TCAATCGTA	TCGAAAAGAA	ACTGAATAAA	CTTGATGACG	TTAATGCCCA	AGTGAATTTA	6480
	ACTACAGAGA	AAGCAACTGT	TGAGTATAAC	CCTGATCAAC	ATGATGTCCA	AGAATTTATT	6540
35	AATACGATTC	AACATTTAGG	TTACGGTGTC	GCTGTAGAAA	CTGTGCAATT	AGACATTACA	6600
	GGTATGACTT	GTGCTGCATG	CTCAAGCCGT	ATTGAAAAAG	TGTTAAATAA	AATGGACGGC	6660
40	GTTCAAAATG	CAACGGTCAA	TTAACAACA	GAGCAAGCTA	AAGTTGACTA	TTATCCTGAA	6720
	GAAACAGATG	CTGATAAACT	TGTCACTCGC	ATTCAAAAAT	TAGGTTATGA	CGCGTCTATT	6780
	AAAGATAACA	ATAAAGATCA	AACGTCACGC	AAAGCTGAAG	CGCTACAACA	TAAATTGATT	6840
45	AAGCTTATCA	TATCAGCAGT	ATTATCTTTA	CCACTATTAA	TGTTAATGTT	TGTACATCTT	6900
	TTCAATATGC	ATATACCAGC	ACTATTTACG	AATCCATGGT	TCCAATTTAT	TTTAGCTACA	6960
	CCTGTACAAT	TTATTATTGG	ATGGCAATTT	TATGTAGGTG	CTTATAAAAA	CTTAAGAAAT	7020
50	GGTGGCGCCA	ATATGGATGT	ACTTGTGCT	GTTGGTACAA	GTGCAGCATA	TTTTTACAGT	7080
	ATTTATGAAA	TGGTTCGTTG	GCTAAATGGC	TCAACAACGC	AACCGCATTT	ATACTTTGAA	7140
55	ACAAGCGCCG	TACTAATTAC	CTTAATCTTA	TTCGGTAAGT	ATTTAGAAGC	TAGAGCGAAG	7200

	TTAAAAGATG	GTAATGAAGT	GATGATTCTT	CTAAATGAAG	TACATGTTGG	AGATACACTT	7320
	ATCGTTAAAC	CAGGTGAAAA	GATACCTGTT	GATGGCAAAA	TTATTAAAGG	TATGACTGCC	7380
5	ATCGACGAAT	CTATGTTAAC	AGGTGAATCT	ATCCCTGTTG	AGAAGAATGT	TGATGATACT	7440
	GTAATTGGTT	CAACGATGAA	CAAAAACGGT	ACTATTACTA	TGACAGCAAC	AAAAGTTGGC	7500
	GGGGACACTG	CGTTGGCAAA	TATTATTAAA	GTTGTGCAAG	AAGCTCAAAG	TTCTAAAGCG	7560
10	CCGATTCAAC	GATTGGCAGA	TATTATTTCT	GGTTATTTCT	TTCTATCGT	TGTTGGTATC	7620
	GCACTATTAA	CATTTATCGT	GTGGATTACT	TTAGTTACAC	CAGGTACATT	TGAACCTGCA	7680
	CTTGTTGCGA	GTATTTCCGT	TCTCGTCATT	GCTTGTCAT	GCGCATTGGG	ACTTGCTACA	7740
15	CCAACTTCTA	TTATGGTAGG	TACTGGTCGC	GCTGCTGAAA	ATGGTATTTT	ATTTAAAGGT	7800
	GGCGAGTTTG	TTGAACGCAC	ACATCAAATT	GATACCATCG	TTTTAGATAA	GACGGGTACC	7860
20	ATTACAAATG	GTCGTCCAGT	CGTGACAGAT	TATCATGGTG	ACAATCAAAC	GCTACAACTA	7920
	CTTGCTACTG	CTGAAAAAGA	TTCTGAACAC	CCATTGGCAG	AAGCCATTGT	CAATTATGCA	7980
	AAAGAAAAGC	AATTAATATT	AACTGAGACA	ACAACATTTA	AAGCAGTACC	TGGCCATGGT	8040
25	ATTGAAGCAA	CGATTGATCA	TCACCATATA	TTGGTTGGTA	ACCGTAAATT	AATGGCTGAC	8100
	AATGATATTA	GCTTGCCTAA	GCATATTTCT	GATGATTTAA	CACATTATGA	ACGAGATGGT	8160
	AAAAC TGCTA	TGCTCATTGC	TGTTAATTAT	TCATTAACTG	GTATCATCGC	AGTGGCAGAT	8220
30	ACTGTCAAAG	ATCATGCCAA	AGATGCTATA	AAACAATTGC	ATGATATGGG	CATTGAAGTT	8280
	GCCATGT TAA	CTGGCGATAA	TAAAAACACT	GCTCAAGCCA	TTGCAAAACA	AGTAGGCATA	8340
35	GATACTGTTA	TTGCAGATAT	TTTACCAGAA	GAAAAAGCTG	CACAAATTGC	GAAACTACAG	8400
	CAACAAGGTA	AGAAGGTTGC	GATGGTTGGT	GACGGTGTA	ATGATGCACC	TGCATTAGTT	8460
	AAAGCTGATA	TCGGTATCGC	CATTGGTACA	GGTACAGAAG	TTGCCATTGA	AGCAGCTGAT	8520
40	ATTACTATT C	TTGGTGCGA	CTTGATGCTT	ATTCCTAAAG	CCATTTATGC	AAGTAAAGCA	8580
	ACCATTTCGTA	ATATTCGTCA	AAATCTATTT	TGGGCATTTC	GCTATAATAT	TGCCGGTATC	8640
	CCTATAGCTG	CATTGGGCTT	ACTTGCGCCA	TGGGTTGCTG	GTGCTGCAAT	GGCACTAAGT	8700
45	TCAGTAAGTG	TTGTCACAAA	CGCACTTAGA	TTGAAAAAGA	TGCGATTAGA	ACCACGCCGT	8760
	AAAGATGCCT	AGATTCCTTA	ATAATGAAGG	ATTCGTTGGT	GATTCTGAGA	TAGGCTAGTG	8820
50	ATTGGCTCTA	TAATGTCGCG	GTTTAYaGt	GGATCTTCGC	TCCAAC TGCA	TATATAGTnA	8880
	CACTTTTTCGC	TTGGCGAATT	AGTGTATCTT	ACCTAATAGc	TCCGCCTATT	AGGTTCATC	8940
	ATTATTATAA	ATAATAAGTA	CACTACGGtT	TACAGTTGGA	TCTTCGCTCC	AACTGCATAA	9000
55							

	GAAATTTTAA ATGTTGAAGG TATGAGCTGT GGTCACCTGCA AAAGTGCTGT TGAATCTGCA	9120
	TTAAATAATA TTGACGGTGT CACTTCAGCT GACGTTAACC TTGAAAATGG TCAAGTAAGT	9180
5	GTTCAATATG ATGACAGTAA AGTTGCTGTA TCTCAAATGA AAGACGCAAT TGAAGATCAA	9240
	GGTTACGATG TCGTTTAATT AGGCAATATT CAACGTCATC AACACCAAAT TAAAAAATCG	9300
10	AACCTAGTAT CTCCAGTCTG CAATACATCT AATGTTGCAT CTAATGCATC GACAATTAGA	9420
	TTTTTAACTG CAGCTTCAGT ATAAAACGCA ATATGTGGTG TTAATATGAC ATCTTCCCTG	9480
15	TCAATCAACG ATTCTAACAA TGGATCGTTC AGTGTTTTGC CCCTTTGATC ACTTGGGAAA	9540
	AGTTTGCGTT CAAATTCATA CGTATCAAGT GCTGCACCTT TAATCACACC ATTGTCTAAT	9600
	GCGTCTAATA ACGCCTTAGT ATCTACTAAA GAACCTCTCG CACAATTGAC AAATACTGCG	9660
20	CCCTTTTAA AATGTTTAAA TAATTCAGCA TTAAATAGAT AATGATTATA TTTCGTTGCA	9720
	GGTACATGTA ATGTCACGAT ATCAGCACCT TCAACCGCTT CCTCAATCGT ATCTTTGTAA	9780
	TCGACATACG TTGCAATTTT AGCATTAGGA AACGGTCGTA TCGGACCACA TCACTTTGAT	9840
25	AACCATTGGC AAATATATCG GCTACTACAC GGCCAATTCTG ACCTGTACCA ATAACAGCTA	9900
	CTTTTAAATC TTTAATGGAT TTCGATAAAA TAGTAGGTTC CCATCTAAAA TCATGCTCCC	9960
30	GCACTTTCGT TTGAATTTGA TTTAAATGAC GAACCACATT AATAGCCTGG TTCACAGCAA	10020
	ACTCCGCAAT TGAATTCGGA GAGTATGACG GCACATTTGA CACAATAAAG TTATACTTGT	10080
	TTGCTAACTC CAAATCATAT GTATCAAATC CAGCACTACG TTGTGCGATT TGTTTAATAC	10140
35	CTAGTTCATT TAATCGTTTA TAAACATGCT CTGATAATGG TATTTGTGTG GATAGCGATA	10200
	AGCCATCATA ACCAGCGACA CCTTCAACAT TGTATCAGT TAATGCTTCT TTAGTAATAT	10260
	CTACCTCAAC ATGATGTTTC TCTGCCCACG CCTTGATATA AGGCATATCT TCATCACGTA	10320
40	CACTCATGAT TTTAATTTT GTCATTTTAA CATCACCTT AACTTTATTA TTCATATAAA	10380
	TATGCTAGTT CTGTTAATCT TATTGCAGCT TCGTCTAATT TCTGGTCATC TAACGCCAAT	10440
45	GAAATCTCA CATAACGATT ACCATTCTCT CCAAATGGTT TCCCTGGAGC AACAAGTATT	10500
	GACTTCTCTT GCACTAAAAA TTGCTCAAAT TGCTCGCTGT CATAACCAGG CGGTGTTTCC	10560
	AACCATACAT ATATGCCACC TTTAGCATGA ACAAATGGCA AATCAGCTTT TGCAAGCATG	10620
50	GCTTCGAATC GGTACGACG TGTTTTAAAT ACATTGCTTT GTTCTTCTAA AAAATCATCA	10680
	TAATGATTCA AAGCATATAT TCGGGCATCT TGTAATGCAC CAAACATCCC AGCATTTGTG	10740
55	TGCGTTTGGT ACTTTTCAA AGCTTGAATC ATATCTTTAT TACCAACTGC AAAACCGACT	10800



	CCATTTTCCG AAGCAAGTAT ACTAGGATTT TTAGCGTCGA AACCGAAAGC ACCATAAGCA	10920
	AAATCATGCA CGATTTTAGT GTCTGTACCT TTAAATTTAG CTATCGCTTC ATCAAAAAC	10980
5	TCTTTCGTAG CTGTCGATCC AGTTGGATTA TTTGGATACG TTAAATAAAT GAGTTTTGTT	11040
	TTATCTATTA TTTGTGAATC AACTTTGGAC CAATCTGGCA AATAATGTGG CGGTTCTAAA	11100
10	TTAAGCGGGA CTGGCTTGCC ATCAGCTAAA AGTACACCTG CTAAATAATC CGTGTAGCCT	11160
	GGATCAGGTA GTAATACATA GTCTCCTGGA TTGATAACAC ATGTTGGTAC TGCCACTAAT	11220
	CCATTTTTTG TACCATATAA AATGCATACT TCATCTTCTT TATCTAACGT CACATTATAT	11280
15	TGTCTTTGAT AAAAATCTAC AATAGCTTGC TTGAACGCTT CTTTACCATG AAAAGCACCA	11340
	TATTTTTGAT TTTCAGGAAT AGTTAGTGCT TTTTGAAAAT GATCAATAAT ACCTTGTGGC	11400
	GTGGGCCCAT CAGGGATTCC AACTGCCATA TTAATTAATG GCAATGGTCC ATGTTTCGATT	11460
20	TTACGTCCCA TCGTTTTCCC GAAATAACTA TCAGGGATAT TTGCTAATTT GTTAGAGATC	11520
	ATCAAATTCC TCCTCTATCA TTAAACATAG CCTGGGCGAC TATCATAATC CTAACAACTT	11580
25	GTATCACTCT CATTTAGATG GTTACAATGA CATCGCCATT CACCGTTATG TTCAACAGAA	11640
	CTTATGACAC ACGTTGTATT GAATGAATTT ATTTTCATTT TAGGTAGGTA TAATATTATT	11700
	GTCAATATTA GGAATTTTCA GATTAATATG CACTCAATCG TTATGATTTA ACTGTCATGC	11760
30	ATATCCGCAT GCGCAACCAG TTAGATATGC TTATATAAAG TATAACGCCC ATCAAGGTAC	11820
	GTATTCAAAC GTGAACCTTA ACAGGCGTCA TTCATTGTGA AATAAACTT CTTAAGCACA	11880
	TACTTATTTT ACTATGCCTT TTACGTTCCC CTTATACTTT TCTCACATCT TTCTCTTAGA	11940
35	CTACTCCCTT ATACGCCCCG CTCAATATCT TTAATCATTT CATCTACAGT TATTTTCGCA	12000
	CTCGTTAAGA CAATAGGAAC GCCTGCACCT GGATGCGTAC TTGCACCTGC AAAATATAAA	12060
40	TCTTTATAAT CTCGCGATAC ATTTTGTGGA CGATAATAAT TACTTTGCGC TAAAGTTGGC	12120
	ATTAAACCGA ATGCCGAACC AAATTTTCGCA TGATACGTTT GCTCAAAATC ATTTGGCGTA	12180
	AAGATTGTTT CTGAAACAAT ATGCGATTTT ATATCTTCAA ATACTTCAAT CGTTGCTAAT	12240
45	TTACGATAAA TAATTTCCCT TATTTGTTGC GTCAAAGCTT CATCTGACCA ATCGATTCCG	12300
	CTACCTGTTT TAAGTTCCGG CGTCGGCATT AGCACATAAA TACCAGTTTT GCCTTCTGGC	12360
	GCAAGTGATT TATCAGCGAC CGCTGGTACA TACACATAAA TAGAAGGATC ATATGATAAA	12420
50	CGTCCCTCAA ATATTTCTTC AATATTGCCT CTAAAGTCAT CTGAAAAAAT AACATTATGA	12480
	AGTCTCACTT GATCTGTCAC ATCAATATCT ATACCGATAT ACATTAAAAA TGCTGAACAA	12540
55	GAGTAATCTA AGTCTGCAAT TTTATGTGGT GGATACTTTT TAATAGGTGC AAAATCTGGC	12600

5 ATGTCACCAT TCACTTTTAT CGCATCGGCC CGTTTGAATT TAGGATCAAT AATAATTTGC 12720  
 TCAATTTTCAAG CATTAGTTC AATATTAACG CCTAAGTCTT TATTTAATTG CGCTAGCCCT 12780  
 10 TGAGCCATGC CATAACATACC GCCTTTAATA AAATGCACAC CAAACATCAT TTCAATCATA 12840  
 GGAATAATTG AATATAGTGA CGGGCCTCGT TTTGGATCAA TTCCTATGTA TAACGTTTGA 12900  
 AACGCTAAAA GCTTTTGTAT CTTTTCGTTA TCAATATAAT GTTCAATTAG CTGATCTGCA 12960  
 15 TGATTTAACG TTTTAACTT AGCACCTTGC ACAAGTGACG TCATATTATA AAAGTCACTC 13020  
 GGTTCGCGAT ACGTTCTTTC TAAGAAATAG CGACGTGCAA TTTCATATTT TTTATAAACA 13080  
 TCCGTTAAAA AGGACATAAA ACCATGCGTT GAACCAGGTT CTATACTTTC TAGCATTTCG 13140  
 TGTAATTCAG CTAAATCTGT AGGCACCGTT ATACGATCAT CGTGGTCAAA ATACACATCG 13200  
 TAAATATAAC GTAATTGTCT CAATTCAATA TAATCTTCAT AATTTTTACC ACACGCTGTA 13260  
 20 AAAACATCTT TATAACATC TGGCATCATG ACAATTGTGG GACCCATATC AAATGTAAAG 13320  
 CCGTCTTTCT TTAATTGATT CATACGCCCG CCTACATTAT TATTTTTTTC AAATATCGTC 13380  
 ACTTCATGAC CTTGAGAAGC AATACGGGCT GCCGCTGCTA ATCCTGTGAC ACCTGCACCA 13440  
 25 ATTACTGCAA TCTTCATTAT TCAACCACCT ATATTCTATG ATATTTACTA TTTATTTTCT 13500  
 GAAACAACTT TGCCTTTTTTCTCTTATCCA CAAAAACACG TTCATGTAAT GTATAGTTAG 13560  
 CCTGTCTCAC TTCGTCCAGT ATTTCAATAT ATATACGTGC TGCTAATTCT ATGATTGGTT 13620  
 GTGCTTCAAT ACTAAATACT TTGATTTGAT CCATAACATC TTGAAAATCT TTTTCTGCGA 13680  
 TAGCTGCATA ATATTCATCAT AAGTCAATAT AATGATTATT AACACCATT TGGTACACTT 13740  
 35 CAGCAATATC AACTTCATAT TGCTTTAATC GTTGCTTACT AAAATATATC CGTTCATTGT 13800  
 CAAAATCTTC ACCGACATCT CTTAATATAT TAAAGGGATC CTCTAGAGTC GACCTG 13856

## (2) INFORMATION FOR SEQ ID NO: 32:

- 40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10088 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

50 ATATATAAAT ATAGATTAAG TATATAGATT AATCAACTTT TTTGGAAGAG CAAATCACGC 60  
 AATCAACAAA TAATATAAGA AGTTTTTGCG ATAGTTTTAA AATAGCTGTA ATAGAATACT 120  
 AAATGTGACA AACTTAGAAC TAATATCAAG TGTGATGTT TTGAATATAA AAATGCTAAT 180

55

EP 0 786 519 A2

	ATAATTGGTT AATATATGAG TAATTAGAAA ATAGACAAAG GATGACGATT TATGTATATC	300
	AATATGAAAG ATTATGGGTT AACAGGCATA AACAAAATA AAGATACTCG AGCAATACAA	360
5	CGTGCGTTAA ATCGTGGAAG ATGTAAACCA ACGACAGTTT ATATACCGAA AGGGACGTAT	420
	GATATTTGCA AACCATTAAAC GATATATGGC AATACAACAC TTTTGTTAGA TAATGAAACT	480
	ATTTTACGCC GATGTCATTG TGGTCCTTTA TTAATAAATG GTCGTCGCTT TGGTTTTTaT	540
10	CGTGGTTATA ATGGACACAG TCATATTCAT ATTAAAGGCG GCAAGTTTGA TATGAATGGT	600
	GTATCGTATC CTTATAACAA TACAGCTATG TGCATTGGGC ATGCTGAAGA TATTCAATTA	660
15	ATAGGTGTGA CCATTAAGAA TGTAAGTGTG GGTGATGCAA TTGATGCTTG TGGGATTAAC	720
	GGACTCTATA TTAATAAGCTG TTCATTTGAA GGATTCATAG ACTATAGTGG CGAACCcTTTT	780
	ATTCTGAAGC AATACAATTA GACATTCAAG TACCTGGTGC TTTTCCAAAA TTCGGAACgA	840
20	CAGATGGTAC GATAACGAAA AATGTCATTA TCGAAGATTG TTATTTTGGG CCTTCAGAAT	900
	TGCCCCGAAAT GGGAAAGTTGG AATCGTGCTA TTGGCTCACA TGCAAGTAGA CATAATCGAT	960
	ACTATGAGAA TATTCATATT AGAAATAATA TATTTGAAGA TATACAAGGT TATGCATTAA	1020
25	CTCCCTTGaA GTATAAGAT GCTTTCATTA TTAATAATAA GTTTATTAAC TGTGaGGGTG	1080
	GCATTAGATA TTTAGGAGTT AGAGATGGTA AAAATGCAGC AGATGTGaTG ACAGGaAAAG	1140
30	ACTTAGGTTT CCAAGCAGGC ATAAATATGA ATATAATTGG AAATGAATTT AAAGGATCAA	1200
	TGTCTAAAGA TGCGATACAT GTACGTAATT ATAATAATGT TAAACATAAA GATGTATTAA	1260
	TCGTTGGGAA TACATTCAAT AATTCGACTC AATCAATTCA TTTAGAAGAT ATTGATACAG	1320
35	TGTTTTTAAG TCCTGTTGAA GCGGGTATTC AAGTTACTAC AATCAATGTA GATGAAATAA	1380
	AAAAGTAAAA AGTTTCGCAT GACATTAGGA TTAAGAATAG TAGATAATTT TTGAAAGCGC	1440
	ATTÉATAAAA CGGTATAAAT ATGCTATAAT AAACCCAATT ATCTGATAAA AGGGGTATTT	1500
40	TGACGGTAAT GATAATACAA GATAGACAAC TTTCTATACT CTAATATAGT GAGTTGAAGT	1560
	AGCTTGTCAT AATCATCATG AGGGGGAAAT TTATGGCTTA TTTCAATCAA CATCAATCAA	1620
45	TGATATCGAA AAGGTATTTA ACATTCTTTT CAAAATCAAA GAAAAAGAAA CCGTTTAGTG	1680
	CGGGACAAC TATTGGACTA ATATTAGGTC CATTACTTTT CCTATTAACA TTATTATTCT	1740
	TTCATCCACA AGACTTACCT TGGAAAGGCG TCTATGTTTT AGCGATTACT TTATGGATTG	1800
50	CGACTTGGTG GATTACTGAA GCAATTCCTA TTGCAGCAAC GAGCTTATTA CCAATTGTGT	1860
	TATTACCATT AGGTCAATATA CTTACACCAG AACAAAGTATC ATCCGAATAT GGCAATGATA	1920
55	TTATCTTTTT GTTTTtagGT GGATTTATTT TGGCAATTGC AATGGAAAGA TGGAATTTAC	1980

	TTGGATTCAT	GGTGGCAACA	GGATTCTTAT	CTATGTTTGT	ATCGAACACT	GCAGCTGTAA	2100
	TGATTATGAT	TCCGATTGGT	TTAGCAATTA	TTAAGGAAGC	ACATGATTTA	CAAGAAGCCA	2160
5	ATACGAATCA	AACAAGTATT	CAAAAGTTTG	AAAAATCTCT	AGTTTTAGCA	ATTGGCTATG	2220
	CAGGTACGAT	TGGTGGCTTG	GGTACATTAA	TCGGAACCCC	GCCATTAATT	ATTTTAAAAG	2280
	GACAATACAT	GCAACATTTT	GGACATGAAA	TTAGTTTTGC	TAAATGGATG	ATTGTAGGGA	2340
10	TTCCAACGGT	CATTGTTTTG	TTAGGTATTA	CTTGGCTCTA	TTTAAGATAT	GTTGCGTTTA	2400
	GACATGATTT	GAAATATTa	CCTGGTGGTC	AGACGTTAAT	TAAACAAAAG	TTAGACGAGC	2460
15	TTGGCAAAAT	GAAGTATGAA	GAAAAGGTAG	TACAAACTAT	CTTTGTACTT	GCTAGCTTAT	2520
	TATGGATTAC	AAGAGAGTTT	CTTCTGAAAA	AATGGGAAGT	TACGTCATCT	GTTGCAGATG	2580
	GTACGATTGC	TATTTTTATA	TCAATATTAT	TATTTATTAT	TCCAGCTAAA	AATACTGAAA	2640
20	AACATCGCCG	TATCATTGAC	TGGGAAGTTG	CAAAAGAGCT	CCCTTGGGGT	GTATTAATTT	2700
	TATTTGGTGG	CGGTTTAGCA	TTAGCGAAAG	GTATTTCTGA	AAGTGGTTTA	GCAAAATGGT	2760
	TAGGCGAACA	GTTGAAATCA	TTAAATGGTG	TTAGTCCGAT	TCTTATTGTA	ATTGTCATAA	2820
25	CAATCTTTGT	CTTATTTTTA	ACTGAAGTGA	CATCTAATAC	TGCAACTGCA	ACGATGATTT	2880
	TACCGATTTT	AGCAACGTTG	TCTGTTGCTG	TTGGAGTGCA	TCCATTACTA	CTTATGGCAC	2940
	CTGCAGCTAT	GGCGGCTAAC	TGTGCATACA	TGTTACCAGT	AGGGACACCA	CCGAATGCAA	3000
30	TTATCTTTGG	TTCTGGTAAA	ATATCTATCA	AACAAATGGC	ATCAGTAGGA	TTCTGGGTAA	3060
	ACTTAATCAG	TGCAATAATT	ATTATTTTAG	TCGTGTATTA	TGTAATGCCT	ATAGTTTTAG	3120
35	GTATTGATAT	AAATCAACCA	CTGCCATTGA	AATAGTAATT	GCAGATTAGA	ACGAAAAATA	3180
	AAAGGTTACA	TTAGCAATTG	CTTGGACGAG	TGGTAACGAA	ACGTATACCG	CAGCATCGTG	3240
	TAAATAACAAT	ACAAACAAAA	GAAAGTCAAC	CAAGGATGGA	TTCCTATTTT	AATCCTTGGT	3300
40	TGACTCTTTA	TTTTATTTAA	ATTGTAGAAC	CTAGAAAATA	AAGTTTAATT	AAAAGCACCA	3360
	ATCATTTCTA	CTTTGAAATC	TAAGGTTTCT	AAAATAGCAA	TGACTTTCTT	TATATCGGTT	3420
	GTAATTGCAG	AATCAGCCTG	AACGAAAAAT	CGATACATAC	CTAATTGTGT	TTTTAAAGGA	3480
45	CGAGACTCAA	TCCAGGATAA	ATTAATATTA	AACAAAGCAA	ATGTATTAAG	CACACTTGCT	3540
	AACAACCCAG	GTTTATCATG	CATTGGTGTA	ATTAAAAACA	TCAATGATGT	CGCATTTTGA	3600
50	TCAAATTGCT	GCTGATTTTT	TATAACTAAA	AAACGTGTCA	CGTTATGTGG	ATAGTCTTCA	3660
	ATATGTGTAT	CAATAGGTGT	AAAACCATAA	GctTCGCCAC	TACCTAAAGG	TGCAATTGCT	3720
	GCAACGCCAT	TTTCAATTTT	AGTCAAACCT	TGAATTGTAC	TGTCGACATA	ATCATAGTCA	3780
55							

	TTTTTAATAT CAGAAATGGA ATCTGTTCCA TTACCATATA ATGCAAAGTT AATATCTAAA	3900
	CGTATTTTAC CGTGTGCAAA GACATCTTGC TGTGCAAGTG CATCTGCCAC AATGTTGATT	3960
5	GTTCCTTCTA TAGAATTTTC AATAGGGACA ACACCAATCG ATGTGTCATC ATCTGCAACT	4020
	GCCTTGATGA CTTCAAATAA ATTTGACTTT GGTGAAAAG TTGCTTCATT TTCAGAAAAA	4080
10	TACTGACGAC AAGCCAAATA TGAAAATGTA CCTTTAGGGC CTAAATAATA TAATTGCATA	4140
	TGCTACACCT CTAATACTT AATGATGGAA AGGGCACTGG TTAGCATTTG ATTCTTTCTT	4200
	TTTATAGAAA AAGTTTGGAT CTTTACTGT ATTGTCATAT CCGTGATGAT AATTTGACGT	4260
15	CAATGTTGGA GATAATGGCG GTGCTAGCCA AGACCATTTT CCGGTAAGTT GACGACCTTG	4320
	TTGTGCTTCG TTACGTTTGA ATAGTTCGAA TTGCTTTGCA GCGGTCAAAT GATCGACAAT	4380
	TGATACGCCT TCTTTTTTAA AGGAATGATA CACAGCATAG TTCAATTCAA CAAGTGCTCG	4440
20	ATCTTTATTA AATGAATTAT TTTTAAAGTG ATCAAATTCA AACGCATCTG CAACTTTTTTC	4500
	TAGTAAATTG TAACGGTAAT CATCAATAAA GTTACGTACG CCAATTTTCA TTACCATATA	4560
	CCAACCGTTA AAGGGTGCAG TTGGATATAC AATGCCACCG ATTTTAAAGT CCATATTGGA	4620
25	AATGATAGGG ACTGCATACC ATTTTAAAGT CAATTTTCTT AATTTTGGAT AATGATTATG	4680
	TTCAATAGGT ACTTCTTTAA TTAATGAAGT AGGATAITCG TAAAATTTAA CTGACTCAIT	4740
30	AGGTAATTGG TAAATCAGTG GTAACACGTC AAAATTAGTA CCTTTTCCTT TCCAACCTAA	4800
	GTGATTGCT AAGCGTGTA CTTCTTTTTT AGCAGGATCA CCACAATTGT CATAGCCAGC	4860
	ATAGCGAATT AATTGATTGT TGAAAATTTT AGGTCCATCC TTTGGAGCAT ATATAGTAAT	4920
35	ATACGGCTTT AATTTACCTT CATTTGTAGC CTGTGTAATA TGATAAGTAA TTGATGATAA	4980
	GAACGATGCT TCGTCAGTAA CATCTCTTGC ATCAATGACA TTTAACGAAT CCCAAAATAA	5040
	ACGACCAATG CAACGATTG AATTACGCCA AGCCATTTTA GCACCATAAA TAAGTTCTTC	5100
40	TTCTGTATGT GTATATGTCC CAGTTTCTTT TATTTCTAGT TCAATGTCAT GTAAACGTTT	5160
	ATTGATAATT TGCCTTTCAT AATGACACTC TTTATACATG TTTTCTATGA AAGCTTGAGC	5220
45	CTCTTTAAAT AACATTAACA ACACCTCGCT TTATATTATA GTCTACATTA TTAAAATACT	5280
	CTTAAAAATT ATGTATATGT CATTAAATTG TTGGTTGATT TTAATTAAAA GTATGGAAAT	5340
	TAAGGGGCTC TTATGTATAT AAAAAATGA ATTATGATAA AATGTAAGAA AATATTTAGG	5400
50	TCGATTGGAG AGATACAAGT GTACCAATTA GAAGACGACA GTTTAATGTT ACATAATGAC	5460
	TTATATCAAA TAAATATGGC TGAAAGTTAT TGGAATGATA ATATTCATGA AAAAAATGGCT	5520
55	GTATTTGATT TGTATTTTAG AAAAAATGCCA TTTAATAGTG GCTATGCTGT TTTAATGGT	5580

	TTAAAGTCTA TTGGCTACAA GGATGATTTT TTATCATATT TAAAAGATTT AAAATTCACA	5700
5	GGCAGCATCC GTTCGATGCA AGAAGGCGAA TTATGCTTTG GTAACGAACC ATTGTTACGC	5760
	GTAGAAGCAC CATTGATTCA AGCGCAATTA ATAGAAACAA TTTTATTAAA CATTGTAAAT	5820
	TTCCATACAT TAATTACAAC AAAGGCTAGC AGAATTCGTC AAATTGCATC AAATGATAAA	5880
10	TTAATGGAGT TTGGTACACG TCGTGCGCAA GAAATTGATG CAGCATTGTG GGGCGCTAGA	5940
	GCTGCTTACA TCGGGGGCTT TGATTCTACA AGTAATGTTA GGGCGGGGAA ATTATTTGGT	6000
	ATACCTGTGT CTGGTACACA TGCACATGCA TTTGTCCAAA CTTATGGAGA CGAATATGTT	6060
15	GCCTTCAAAA AATATGCTGA AAGACATAAA AATTGTGTGT TCCTAGTAGA TACATTCCAT	6120
	ACTTTAAAAT CTGGCGTGCC AAATGCAATA AAAGTTGCAA AAGAATTAGG TGACAAAATT	6180
	AACTTTGTAG GTATTGATT AGATTCTGGA GATATCGCTT ATTTATCTAA AGAGGCAAGA	6240
20	CGTATGCTTG ATGAAGCAGG ATTTACTGAA ACTAAAATTA TCGCGTCTAA TGATTTGGAT	6300
	GAAGAAACGA TTACGAGTTT GAAAGCACAA GGTGCAAAAG TAGATTCTTG GGGCGTTGGT	6360
25	ACAAAGCTGA TTACAGGATA CGATCAACCA GCATTAGGTG CAGTATATAA ACTTGTAGCT	6420
	ATTGAAAATG AAGATGGTTC ATATAGTGAT CGTATTAAAT TATCAAATAA CGCTGAAAAG	6480
	GTTACGACGC CAGGTAAGAA AAATGTATAT CGCATTATAA ACAAGAAAAC AGGTAAGGCA	6540
30	GAAGGCGATT ATATTACTTT GGAAAATGAA AATCCATACG ATGAACAACC TTTAAAATTA	6600
	TTCCATCCAG TGCATACTTA TAAAATGAAA TTTATAAAAT CTTTCGAAGC CATTGATTTG	6660
	CATCATAATA TTTATGAAAA TGGTAAATTA GTATATCAAA TGCCAACAGA AGATGAATCA	6720
35	CGTGAATATT TAGCACTAGG ATTACAATCT ATTTGGGATG AAAATAAGCG TTTCTGAAT	6780
	CCACAAGAAT ATCCAGTCGA TTTAAGCAAG GCATGTTGGG ATAATAAACA TAAACGTATT	6840
	TTTGAAGTTG CGGAACACGT TAAGGAGATG GAAGAAGATA ATGAGTAAAT TACAAGACGT	6900
40	TATTGTACAA GAAATGAAAG TGAAAAAGCG TATCGATAGT GCTGAAGAAA TTATGGAATT	6960
	AAAGCAATTT ATAAAAAATT ATGTACAATC ACATTCAITT ATAAAATCTT TAGTGTTAGG	7020
45	TATTTCAGGA GGACAGGATT CTACATTAGT TGGAAACTA GTACAAATGT CTGTTAACGA	7080
	ATTACGTGAA GAAGGCATTG ATTGTACGTT TATTGCAGTT AAATTACCTT ATGGAGTTCA	7140
	AAAAGATGCT GATGAAGTTG AGCAAGCTTT GCGATTCATT GAACCAGATG AAATAGTAAC	7200
50	AGTCAATATT AAGCCTGCAG TTGATCAAAG TGTGCAATCA TTAAAAGAAG CCGGTATTGT	7260
	TCTTACAGAT TTCCAAAAAG GAAATGAAAA AGCGCGTGAA CGTATGAAAG TACAATTTTC	7320
55	AATTGCTTCA AACCGACAAG GTATTGTAGT AGGAACAGAT CATTGAGCTG AAAATATAAC	7380

	TAAACGACAA GGTCTGCAAT TATTAGCGTA TCTTGGTGCG CCAAAGGAAT TATATGAAAA	7500
	AACGCCAACT GCTGATTTAG AAGATGATAA ACCACAGCTT CCAGATGAAG ATGCATTAGG	7560
5	TGTAACCTTAT GAGGCGATTG ATAATTATTT AGAAGGTAAG CCAGTTACGC CAGAAGAACA	7620
	AAAAGTAATT GAAAATCATT ATATACGAAA TGCACACAAA CGTGAACTTG CATATACAAG	7680
10	ATACACGTGG CCAAAATCCT AATTTAATTT TTTCTTCTAA CGTGTGACTT AAATTAAATA	7740
	TGAGTTAGAA TTAATAACAT TAAACCACAT TCAGCTAGAC TACTTCAGTG TATAAATTGA	7800
	AAGTGATGA ACTAAAGTAA GTATGTTTAT TTGAGAATAA ATTTTATTT ATGACAAATT	7860
15	CGCTATTTAT TTATGAGAGT TTTCGTACTA TATTATATTA ATATGCATTC ATTAAGGTTA	7920
	GGTTGAAGCA GTTTGGTATT TAAAGTGTA TTGAAAGAGA GTGGGGCGCC TTATGTCATT	7980
	CGTAACAGAA AATCCATGGT TAATGGTACT AACTATATTT ATCATTAAAG TTTGTTATGT	8040
20	AACGTTTTTA ACGATGCGAA CAATTTTAAC GTTGAAAGGT TATCGTTATA TTGCTGCATC	8100
	AGTTAGTTTT TTAGAAGTAT TAGTTTATAT CGTTGGTTTA GGTTTGGTTA TGTCTAATTT	8160
	AGACCATATT CAAAATATTA TTGCCTACGC ATTTGGTTTT TCAATAGGTA TCATTGTTGG	8220
25	TATGAAAATA GAAGAAAAC TGGCATTAGG TTATACAGTT GTAAATGTAA CTTCAGCAGA	8280
	ATATGAGTTA GATTTACCGA ATGAACTTCG AAATTTAGGA TATGGCGTTA CGCACTATGC	8340
30	TGCGTTTGGT AGAGATGGTA GTCGTATGGT GATGCAAATT TTAACACCAA GAAAATATGA	8400
	ACGTAAATTG ATGGATACGA TAAAAAATTT AGATCCGAAA GCATTTATCA TTGCGTATGA	8460
	ACCTCGAAAC ATACATGGTG GATTCTGGAC TAAAGGCATT CGTCGTAGAA AGCTTAAAGA	8520
35	TTATGAACCA GAAGAACTGG AAaGTGTAGT AGAaCATGAA aTTCmAGTA AaTGAGaATG	8580
	AAmCAATtGC TGATTGTTTG TCACGAATGA AAtGCAAGGG TATATGCCGG TAAAACGTAT	8640
	TGAaAAACCC GTGTTTCAAG AGCAAAAAGA TGGCACGGTT GAAGTATCAC ATCAAGAAAT	8700
40	CGTTTTTGTA GGTAAGAAAA TCCAATAACA TAATCCAATT TAAATAAAGA CTATTTGAAG	8760
	AGGAAAGGCT ATTCAAAGTT TGAGTAATTT TACTTTGAAT AGCCTATTTG TTTATACATG	8820
	CAAGATGCTC GATCCATATT GTATGAGAAA CCCCCAGCAA GCTATATAAA GCATATGCTG	8880
45	GGGGTTCTTA ATATTTTAAA AATTATTGTT AGATTATATA TATCGTCGCT TTTTCTAAAA	8940
	CAATCTCATC GCATGAAATT TTTTCTTCCT AGAGACCTTT AATAAGATTA ATAGTTTACT	9000
50	TAATCATATC TAGATAGTCT TATGACTTAT GCTTAATGAA AGTCATTCTA GGAGAAGTTC	9060
	CCAAAGCTTC TGTGTTTATA ATTGTTAGTA GTATTTTATT ATCATTTGGT ATAAATATTT	9120
	CAATAACAAT TGAGCTATTA TTTTATTAT ATAATGTGAG TTGTTTGTGT TCTGTATTTA	9180
55		

CATTTAAATC TTGAGGATGC CATTCTCCCT CAATAATATT AAGATAATAC TTAGCCTCTG 9300  
 AATTACATTT GAATTTATCA ATACTAAATA ATTCAATTTG TTCCATAATA TTATTTACCT 9360  
 5 TTCTAAAATA CAAATTTTAA TAACCATAAA TAGATGAATA CCATCGATAA TGGTCGCCAT 9420  
 TGGATACTGG AATAACATTG TTTTTCAGCAT CTTGAGTCAT AAAACCATTG TCCCATGGAT 9480  
 TCCATATAAT TATAACCTCT TGTCCATTAT CTAATTTAGC GTTCCCAACA ACTGCCATGG 9540  
 10 CATGCCCTGC GTGCATACCA TTTCTTGATT CTAATTTAGC GTTCCCAACA ACTGCCATGG 9600  
 TATTATTTTT AGTAAGATTG TCAACTTCAT TATATGTAGT CATTCTATTA AGAAGTTGTG 9660  
 15 GACTTCTTCC CTGAGTTTGT CCAAAATAAA TCATCTCTCT TGGCGTTAAA CCAGTAAATT 9720  
 GGAATCGTTG TCCTTGTAAG TTTGGGTGTA AAAATCTCAT CACAGCTTCT GCATGATATT 9780  
 TGTAGTATT ATAAGTCGCA TTTAGTAATT CAGACATCGT ATAGCCTGCA CACCAACCAT 9840  
 20 TGTACCTTG AGTTTCTCTT ATCTTGAAAT TCTCAAGTTT ATTTATATAT TGsTCGTTGT 9900  
 AAGTATAATT ATTACTTTTA AATTGACTAG TTGGCATAGT GACAGAAGCT TTTTGCTTTA 9960  
 GTTGCCTTAC ATTATTGCCA GTAGGTATAC TCTCAGTCTT TnTnAACTnT nTATCTTCTA 10020  
 25 GACGTGGTGT TTTAGTACT AGTTTAGCTT TATGATTTTG AGTACCACAT AGTAACCTTT 10080  
 TGAGTTGT 10088

## (2) INFORMATION FOR SEQ ID NO: 33:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7563 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

40 CGGAAACGnA CCCnATGCGT ATGCTTGACG TGCCAAAATT AAATACGAAG TTCATAGCTT 60  
 TGAGGTACCA GAAGAACATT TATCTGGTCA AGAAGTCGCA GnACTCATAC AAGCAAATGT 120  
 TAAAACAGTA TTTAAAACGC TTGTTCTAGA AAATACAAA CATGAACATT TTGTATTTGT 180  
 45 TATCCAGTA AGTGAACTT TAGATATGAA AAAGGCAGCT GCTTTGGTTG GAGAGAAGAA 240  
 ATTGCAGCTT ATGCCTTTAG ATAATTTGAA AAATGTAACG GGATACATTC GTGGTGGGTG 300  
 TTCGCCTGTT GGTATGAAAA CATTGTTTCC AACAGTCGTT GACAAATCGT GTGAAAATTA 360  
 50 TAGTCATATC AGTGTGAGTG GTGGGCTTCG AACAAATGCAA ATCACAATAG CTGTTGAGGA 420  
 TTTGATTACA ATAATAAAG GCAAAATTGG AGCAGTTATC CATGAATGAT TAATAACAAC 480



	TGCCCACTC CTTTTGATT GAATTAGCAT TTTACGATCA TAAACAGTCA TTATAATTGA	600
	GTATTTGAAC ATAAAAATGT AATTTTATCG TAACAATTG AGTGTGTTG ATTGTTTTG	660
5	GTAATTTATG ATTGAAAAGT GAAAGCGTAC TCATTATAAT ACAAAGTGAG ATGGGGTGAT	720
	GATGATAATT ACTGaAAAA GACACGAGTT AATATTAGAA GAACTTTCGC ACAAAGATTT	780
	TTTGACTTTA CAAGAATTAA TAGATCGAAC TGGTTGCAGT GCTTCAACAA TACGArGAGA	840
10	TTTATCTAAA CTACAACAAT TAGGGAAATT GCAACGTGTG CATGGTGGTG CAATGTTAAA	900
	AGAAAATCGT ATGGTTGAGG CGAATTTAAC TGAAAAATTA GCAACGAATC TTGATGAAAA	960
15	GAAAATGATT GCTAAAATAG CAGCTAATCA AATCAACGAT AATGAATGCT TATTTATCGA	1020
	TGCTGGTTCA TCTACATTGG AGCTAATTAA ATATATTCAA GCGAAAGATA TCATTGTGGT	1080
	AACCAATGGT TTAACACATG TAGAAGCTTT ACTTAAAAA GGTATTAAAA CAATTATGCT	1140
20	AGGTGGTCAA GTTAAAGAAA ATACACTTGC TACGATTGGT TCTAGTGCTA TGGAGATATT	1200
	AAGACGATAT TGTTCGATA AAGCTTTTAT CGGGATGAAT GGATTAGATA TTGAACTTGG	1260
	ATTAACACT CCCGATGAGC AAGAGGCATT AGTTAAACAA ACAGCAATGT CATTAGCCAA	1320
25	TCAATCATTT GTACTTATAG ATCATTCTAA GTTTAATAAA GTATATTTTG CTCGTGTACC	1380
	TTTGCTAGAA AGTACGACAA TCATCACATC TGAAAAAGCA TTAAATCAAG AATCGTTAAA	1440
30	AGAATACCAA CAAAAGTATC ACTTTATAGG AGGGACTTTA TGATTTATAC AGTGACTTTC	1500
	AATCCTTCAA TTGACTATGT CATTTTACG AATGATTTTA AAATTGATGG TTTGAACAGA	1560
	GCAACAGCAA CATATAAATT CGCTGGGGGG AAAGGTATTA ATGTCTCGCG CGTCTTAAAG	1620
35	ACATTGGATG TTGAGTCAAC TGCCTTGGA TTTGCAGGTG GATTTCTGG GAAATTCATT	1680
	ATAGATACAT TAAATAACAG TGCAATTCAA TCGAATTTTA TTGAAGTTGA TGAAGATACA	1740
	CGTAJTAATG TGAAATTTAA AACAGGACAA GAAACAGAAA TCAATGCACC GGGTCCTCAT	1800
40	ATAACGTCAA CACAATTTGA ACAACTGTTA CAACAAATTA AAAATACAAC AAGCGAAGAT	1860
	ATAGTTATTG TTGCTGGAAG TGTACCAAGT AGTATTCCAA GCGATGCGTA TGCGCAAATT	1920
	GCACAAATTA CAGCACAGAC AGGTGCTAAA TTAGTAGTCG ACGCTGAAAA AGAATTGGCT	1980
45	GAAAgCGTTT TACCATATCA TCCACTATTT ATTAAACCTA ATAAAGATGA ATTAGAAGTG	2040
	ATGTTTAATA CAACAGTGAA CTCAGACACA GATGTTATTA AATATGGTCG TTTGTTAGTT	2100
50	GATAAAGGTG CGCAATCTGT TATTGTCTCG CTTGGCGGTG ATGGTGCTAT TTATATTGAT	2160
	AAAGAAATCA GTATTAAAGC AGTTAATCCA CAAGGGAAAG TGGTTAATAC AGTTGGCTCT	2220
	GGTGATAGTA CAGTTGCAGG CATGGTGGCT GGAATTGCTT CAGGTTTAAAC GATTGAAAAA	2280
55		

	CGGGACGCTA TAGAAAAAAT AAAATCACAA GTTACGATTA GCGTACTTGA TGGGGAGTGA	2400
	AAATAATGAG AGTAACAGAG TTATTAACAA AAGATACAAT AGCAATGGAT TTAATGGCAA	2460
5	ATGACAAAAA TGGTGTATT GATGAGTTAG TAAATCAATT AGACAAAGCA GGTAAATTAA	2520
	GTGATGTCGC GTCATTTAAG GAAGCGATTG ACAATCGAGA ATCACAAAGT ACAACTGGTA	2580
	TCGGCGAAGG TATTGCCATT CCACATGCCA AAGTGGCCGC AGTTAAGTCA CCAGCTATTG	2640
10	CGTTTGGTAA ATCTAAAGCA GCGGTAGATT ATCAAAGTTT GGATATGCAA CCAGCACACT	2700
	TATTCTTTAT GATTGcAGcG CCAGAAGGTG GCGCCCAAAC ACATCTAGAT GCTTTAGCTA	2760
15	AGTTGTCTGG TATTTTAATG GATGAAAATG TACGTGAGAA ATTATTACAT GCTTCATCAC	2820
	CTGAAGAAGT ACTAGCGATC ATAGATGAGG CTGATGATGA AGTGACAAAA GAAGAAGAGG	2880
	CAGAAGCTGA AGCACAAACAA GTTGCAACTG CAGAACAATC ATCTAAACAA TCTAATGAGC	2940
20	CATATGTGTT AGCAGTAACT GCTTGTCCAA CAGGTATTGC ACACACATAT ATGGCACGTG	3000
	ATGCATTGAA AAAGCAAGCG GATAAAATGG GTATTAAAT TAAAGTAGAA ACGAATGGTT	3060
	CAAGCGGCAT TAAAAACCAT TTAAGTGAAC AAGATATTGA AAATGCAACA GGTATCATTG	3120
25	TTGCTGCTGA TGTTTCATGTT GAGACGGATC GCTTCGATGG TAAAAATGTC GTAGAAGTAC	3180
	CAGTAGCAGA TGGTATTAAA CGCCAGAAAG AATTAATTAA TAAAGCATTG GATACAAGTC	3240
	GTAAACCTTT TGTTGCCCCG GATGGTCAAA GAAAAGGTAA CTCAAATGAC AGTCAAGAAA	3300
30	AATTAAGCCC AGGTAAAGCA TTCTATAAAC ACTTAATGAA CGGTGTCTTCT AACATGTGTC	3360
	CACTTGTAAT ATCTGGTGGT ATTTTAATGG CAATTGTATT TTTATTTGGA GCAAATTCAT	3420
35	TTAATCCAAA AAGCTCAGAG TACAATGCGT TTGCAGAGCA GCTTTGGAAC ATTGGTAGTA	3480
	AAAGTGCATT CGCGTTAATC ATTCCAATTT TATCTGGATT CATTGCACGT AGTATTGCCG	3540
	ATAAACCTGG TTTCGCTTCA GGTCTTGTAG GTGGTATGTT AGCAATTTCA GGTGGTTCAG	3600
40	GATTTATTGG TGGTATTATT GCAGGTTTCT TAGCAGGTTA CTTAACACAA GGTGTTAAAG	3660
	CCATGACACG TAAGTTACCA CAAGCATTAG AGGGATTAAA GCCAACATTA ATTTATCCAC	3720
	TATTAACAGT GACGGCTACA GGCTTATTGA TGATTTATGC CTTTAATCCA CCAGCATCTT	3780
45	GGTTAAATCA TTTGTTATTA GATGGATTAA ACAATTTATC AGGTTCTAAT ATTGTATTAT	3840
	TAGGTTTAGT TATTGGCGCT ATGATGGCGA TTGATATGGG CGGTCCATTC AACAAAGCGG	3900
50	CATATGTTTT TGCAACAGGT GCGTTGATTG AAGGTAATGC AGCACCATT ACAGCTGCAA	3960
	TGATTGGTGG TATGATTCCA CCGTTAGCAA TTGCACAGC GATGTTAATT TTTAGACGTA	4020
	AATTTACAAA AGAACACGT GGTTCATTA TCCCTAACTA TGTGATGGGT ATGTCATTTA	4080
55		

## EP 0 786 519 A2

	TGATTGGTTC AGGTATAGGT GGC GCAATTG CTTTAGGCTT AGGTTACGA ATTACTGCGC	4200
	CACATGGTGG TATTATTGTA ATTGTTGGTA CTGATGGTGC ACACCTACTT CAAACTCTTA	4260
5	TTGCACTTCT AGTTGGCACA TTAGTTTCAG CATTAATTTA CGGTTTAATC AAACCAAAGT	4320
	TAACTGAAAC AGAAATCGAA GCTTCAAAT CAATGGACGA GTAGTTTAA TGATGTAAAA	4380
	TGATTGTTAG CAAAGAGCTT CATATTAAGT TGTATGTTCA ATGAATATAT GTTAGTTTAA	4440
10	TATATCGTGT TAACGGTAGC TTATACAAAG CTGTAAAAAC ACTTCTATT AATTCAGTTT	4500
	TTATGAATG ATATGAAAGT GTTTTATTT TTAGATAAAT GAATGAAGAA ATAGACACCA	4560
15	CAAATGTATA GACTTTTTTA ATATTTTGCA AAAAGTTATG CCAAACGAAG CAGATATAGT	4620
	AAAATATGAG TGTCTTAAAG TGAAAATTTA TAAATAAAGA AGGGTTTATA CGTGTCAGAA	4680
	TTAATTATAT ATAACGGCAA AGTTTATACT GAAGATGGCA AAATCGATAA TGGTTACATT	4740
20	CATGTGAAAG ATGGACAGAT TGTGCAATT GGAGAAGTGG ATGATAAAGC AGCAATTGAT	4800
	AATGATACGA CAAATAAAAT TCAAGTGATT GATGCTAAAG GTCATCATGT ATTACCAGGT	4860
	TTTATTGATA TACATATTCA TGGTGGTTAT GGTCAAGATG CAATGGATGG GTCATACGAT	4920
25	GGCTTAAAT ATCTATCCGA AAATTTGTTG TCTGAAGGGA CGACATCATA CTTGGCCACT	4980
	ACAATGACGC AATCGACTGA TAAATAGAT AATGCACTTA CAAATATTGC TAAATATGAA	5040
30	GCGGAGCAAG ATGTTCAAA TGCAGCGGAA ATTGTAGGTA TACATTTAGA AGGACCATTT	5100
	ATATCTGAAA ATAAAGTTGG TGCTCAACAT CCGCAATACG TTGTACGCCC ATTTATCGAT	5160
	AAAATTAAAC ATTTTCAAGA GACTGCTAAC GGATTAATAA AGATTATGAC GTTTCACCT	5220
35	GAAATTGAAG GTGCAAAAGA AGCGCTTGAA ACGTATAAAG ATGACATTAT TTTTCAATT	5280
	GGTCATACAG TAGCAACATA CGAAGAAGCA GTTGAAGCTG TTGAGCGAGG AGCTAAACAT	5340
	GTCACGCATT TATATAATGC AGCGACGCCA TTCCAACATA GAGAACCAGG TGTTTTTGGA	5400
40	GCAGCATGGT TGAATGATGC TCTACATACC GAAATGATTG TTGATGGCAC TCATTCTCAT	5460
	CCGGCATCGG TTGCAATTGC TTACCGTATG AAAGGTAATG AACGTTTTTA TTTAATTACC	5520
	GATGCAATGC GTGCAAAAGG TATGCCTGAA GGAGAATATG ATTTGGGTGG AAAAAAGTA	5580
45	ACTGTTCAAT CGCAACAAGC ACGTCTGCA AATGGTGCGC TTGCTGGTAG TATTTTAAAA	5640
	ATGAATCATG GGTTACGTAA CTTAATATCA TTTACAGGTG ATACATTAGA TCATTTATGG	5700
50	CGAGTAACAA GTTTAAATCA AGCCATTGCA TTAGGTATCG ATGATAGAAA AGGTAGTATT	5760
	AAAGTAAATA AGGATGCAGA TCTTGTATT CTAGATGATG ATATGAATGT AAAATCTACA	5820
	ATAAAACAAG GCAAGGTTCA CACATTTAGC TAATAAATAA TCATAATTAA ATGTATGCAA	5880

55

	TTTTCTGGGG GTGTCTAAAT GGGAAGGCGA TAACATGTAG TTGTAATTTA AGTCATAGTG	6000
	ATAAATTTGA ATGCGTGTTA CCCATGAGTG ACACATATAA CATGGAGGTG AATCCCTAGA	6060
5	AATAGGGAAT TAATTGGAAA CTTCGACCAT AATTAGTTTG ATTATATTTA TTCTATTAAAT	6120
	TGCATTAACC ACTGTATTTG TTGGTTCAGA ATTTGCATTA GTAAAAATTA GAGCAACAAG	6180
	AATTGAACAG CTAGCAGATG AAGGAAATAA ACCTGCTAAA ATAGTAAAAA AGATGATTGC	6240
10	TAATCTAGAT TATTATCTTT CTGCTTGTC A GTTAGGTATA ACAGTAACAT CTTTAGGGTT	6300
	AGGTTGGCTT GGTGAACCAA CGTTTGAAAA GCTATTACAC CCAATATTTG AAGCAATCAA	6360
15	TTTACCAACT GCATTAACGA CGACGATTTT GTTTGCAGTG TCATTTATAA TCGTTACGTA	6420
	TTTGCATGTA GTACTTGGTG AATTAGCGCC TAAATCTATA GCTATTCAAC ATACTGAAAA	6480
	GCTTGCTTTA GTATATGCAA GACCATTGTT CTATTTCCGT AACATTATGA AACCATTGAT	6540
20	TTGGCTGATG AATGGTTCTG CACGTGTTAT TATTAGAATG TTTGGTGTAA ATCCTGATGC	6600
	CCTAACTGAT GCAATGTCAG AAGAAGAAAT CAAAATTATT ATTAACAATA GTTATAATGG	6660
	TGGAGAAATC AACCAAACTG AATTGGCATA TATGCAAAAT ATCTTTTCAT TCGATGAAAG	6720
25	ACATGCAAAA GATATAATGG TACCTAGAAC TCAAATGATT AACTAAATG AACCTTTTAA	6780
	TGTAGACGAA TTACTAGAAA CAATAAAAGA ACATCAATTT ACGCGTTATC CAATTACTGA	6840
	TGATGGTGAT AAAGACCACA TTAAAGGATT TATTAACGTC AAAGAATTTT TAACTGAATA	6900
30	CGCTTCTGGA AAAACGATTA AAATAGCAAA CTATATaCAT GAGTTGCCAA TGATTTCAGA	6960
	GACAACACGT ATCAGTGATG CATTAAATTAG AATGCAACGT GAACATGTAC ATATGAGTCT	7020
35	TATTATAGAT GAATATGGTG GAACGGCAGG TATTTTAAAC ATGGAAGATA TTTTAGAAGA	7080
	AATCGTTGGA GAAATTCGTG ATGAATTTGA TGATGATGAA GTGAATGATA TCGTTAAAAAT	7140
	TGATaATAAG ACATTCCAAG TAAATGGCAG AGTACTATTG GATGATTTAA CTGAAGAGTT	7200
40	CGGTATAGAA TTTGATGACT CTGAGGATAT TGATACGATA GGTGGATGGT TACAATCTCG	7260
	TAATACCAAT TTACAAAAAG ATGATTACGT GGATACAACT TATGATCGCT GGGTTGTTTC	7320
	AGAAATCGAT AACCACCAAA TTATTTGGGT GATATTAAAC TATGAATTTA ATGAAGCGAG	7380
45	ACCTACTATC GGACAGTCTG ATGAAGATGA AAAATCAGAA TAGATATTAA TATATAAACC	7440
	AACTAAGAAT GATTTAATTC ATTTTGGTT GGTATTTTTT TTGACTAAAA TTAAnGAAAA	7500
50	GTGAAAATAG TATTGGAAC CAATATCTTT AATGATTTAA TGAATAAnTT TTATTGAAAG	7560
	CGA	7563

(2) INFORMATION FOR SEQ ID NO: 34:

55

(A) LENGTH: 3492 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

10	TTATATCAAC TTCATGGCGG AACCATTGAT GACCCATTAG ACGAAACAAT AAGCGCATT	60
	sATGAATTGA AACAGAAGG AATTATACGT GCTTACGGTA TTTCTTCTAT TCGCCCAAAT	120
	GTAATTGATT ATTATTTAAA ACATAGTCAA ATCGAAACGA TAATGTCTCA ATTCAATTG	180
15	ATTGATAATC GTCCAGAATC ATTATTAGAT GCAATTCACA ACAATGATGT TAAAGTATTG	240
	GCAAGAGGAC CTGTGTCTAA AGGATTATTA ACTTCAAACA GTGTTAATGT GCTCGACAAT	300
	AAATTTAAAG ATGGTATTTT TGATTATTCT CATGATGAAT TGGGTGAAAC AATAGCCTCT	360
20	ATTAAAGAAA TTGAAAGTAA TTTATCTGCA TTGACATTTA GTTATTTAAC ATCACATGAC	420
	GTGCTTGGTT CCATCATTGT AGGTGCAAGT AGCGTCGACC AATTAAAAGA AAATATTGAA	480
	AACTATCATA CTAAAGTTAG TTTAGATCAG ATTAAAACAG CAAGAGCTCG TGTAAGGAT	540
25	TTGGAATATA CCAATCATT	600
	AGTGTAGAAG TCATTTTCAG TAATAAAAAC AGCAGCATGA	660
	GGCGTTTCAT TATAAAAATG CCTTACTGCT GTTGTTTATG TACAATTCGC TATAATTTAT	720
30	GATTATGATT ACTCACTTAT GATAGAAATT AAAGCGTTGT CCTCACGCAT CAGTATTTAG	780
	TAATTTGCGC TTGCGGCATT GCCTTAAGCA AACTTCTGCC ACTTCATCTC TTAATAATTT	840
	TATTAAAACA TCTTTCTATA TTTCACTTCG CATGTTGATT CATCATTATT AGTTATTATT	900
35	TGTACACCCA GCACATTTCC TTGCAACACA AGTAGTTTGA ATTTTTCACA AGTATAATAT	960
	AATGTACCGT CTGAAATTTG GTCTACAGAA ATATCGCCTA AAATATCCAG CACTGTAAAT	1020
	TCTTCAAATA CTGATAGTTG TTCCGCATAT CGTACACAAA GTCTTACCAC ACTCTCCGAT	1080
40	TGACAGTTCA TTGCCATCCC ACCTATTTAT GCTTTATTTT TAAATAATTT AGGGAAACAT	1140
	CGTTCAAAAA ATCTAGGCGC AATTTGATAC ATTTTCAACG CATGaTGCAT CCATTTAGGC	1200
	CGATTAAATTT CCAATTGTTT TGTTTAAATG CCATAAATGA TATCTTCTGC AAGCTGATTA	1260
45	GCATCAAGCA TAATTTCCCC CATCTTTTTA gCATACTTCA TTGATGGGTC GGCTTTTTGA	1320
	TGAAAAGGTG TATCAATCGG GCCAACATTA ACTGTCATGA TATGTAAGTT TGGTGACTCT	1380
50	AGTCTTAAAG CATTCAATTA TGATATAAAC CCTGCTTTCG ATGCCCCATA ATGTGCAGCA	1440
	TTTGCTTG TG GAAAAATGC AGCTTGACTT GAAATACCTA CAATATGTGC GTTAGATGTT	1500
	AAATATGGTC TCAACACAGT ATATAAACA TTAATACTAA TTAAATTAAG CTGATACGTT	

55

	TAAATGAATC CATCGAATGA TGTATTGTCT TCAAATTGCA GTGCCTGTAT CGACTTCAAA	1620
	TCATTTAAGT CACAAGGAAT AACATTTATA GTTTTCCCCA ATTCCTGTTC AAAGATTCTA	1680
5	GTTGCTTTAT CAACATCACG CACCAACAAC GTTACATGCA CTTTATTTTC TAGTAACTTT	1740
	CGGACAATCG ATAAACCTAA ACCACTCGTA CCACCAGTCA CTATAAAATG TTGTCCTTTC	1800
10	ATCAATTAAC CTTCTTTTTC AATTATATAG AATGCAATTT ATCAACTTTA CATAATTGAG	1860
	ACAAGTTGAT TATCTTTTCCT AATATATATA CAATAATAAG AAAATATAAC ATACAAATCA	1920
	AAAATAAAG GGATGTGaCG TTAATGrAAC TCGTATTTTA TGGAGCTGGT AATATGGCAC	1980
15	AAGCTATATT TACAGGrATT ATTAACtMa GCAACTTAGA TGCCAATGAT ATATATTTAA	2040
	CAAATAAATC TAATGAACAA GCTTTAAAAG CATTGCTGA AAACTAGGT GTTAACTATA	2100
	GTTATGAtGA TGCGACATTA TTAAAAGATG CAGAyTATGT ATTTTtagGT ACCAAACCAC	2160
20	ATGACTTTGA TGCTCTAGCA ACACGCATCA AACCACATAT TACAAAAGwC AATTGCTTCA	2220
	TTTCAATTAT GGCAGGTATT CCGATTGATT ATATTAAACA ACAATTAGAA TGCCAAAATC	2280
	CaGTTGCTAG AATTATGCCA AACACAAATG CGCAAGTTGG ACACTCTGTT ACTGGCATTa	2340
25	GTTTTTCAAA CAACTTTGAC CCTAAATCTA AAGATGAAAT TAACGATTTA GTTAAAGCAT	2400
	TTGGTTCTGT AATTGAAGTA TCAGAAGATC ATTTACATCA AGTAACAGCT ATCACCgGAA	2460
30	GCGGCCcAGC ATTTTTATAT CATGTATTcG AGCAATATGT TAAAGCTGGT aCsAAACTTG	2520
	GTCTAGAAAA AGAACAAGTT GAAGAATCTA TACGCAACCT TATTATAGGT ACAAGTAAGA	2580
	TGATTGAACG TTCAGAtTTG AGCATGGCTC AATTAAGAAA AAATATTACC TCTAAAGGTG	2640
35	GTACGACACA AGCTGGCCTT GATACATTGT CACAATATGA TTTAGTATCT ATTTTCGAAG	2700
	ATTGTCTAAA CGCTGCCGTC GACCGTAGTA TTGAACtTTC TAATATAGAA GACCAATAAA	2760
	AACAfACCCG CCAACACATG TATGCATCAT CGCAAGCACT GTGTTTGACG GGTtATTTTT	2820
40	ATAATTTATT GTTATTTGGC AAGCATTGTT TATTACTTTG TCATTAGATT TTAAAACTAT	2880
	CAAAATCTTT TACAAAATTA AAATTAGGTG TATCTTCATT TTGTATCAAT GTTTGATAAA	2940
	TTTCATTTAT ATCTTCTGTA TTATAGCGAT TGCTCAAATG TGTAATCAAC GTACGTTTAA	3000
45	CATTGGCTTC TTTTATCAAT GCAAATACGT CTTCAATATG GCTATGATGA TAATTGTTGG	3060
	CTAAATGCTT TTCACCATCT ATATAGGTCG CTTCATGTAC CATCACATCA GCATCTCTAG	3120
	AAATCACACG TTCATTAGAA CATGGTTTTG TATCACCAAA AATTGCTACA ACTGGACCCT	3180
50	GTTTGGACTC ACCTCTAAAA TCTTTTGATT GATAAACTTG ACCATTATGT TCAAATGTAT	3240
	CATGAGATTT TACTTCTTGA TATTTAGGAC CTGGTTCAAG ACCAATGTTT TTAAACGCTT	3300
55		

CATGATTAAG TAAATGCGCC TCTACAGTAA AACCATCCAT GATGATATGT CAGATGATCA 3420  
 TCGATTTCAA TATATGtAAT TGGATAGTTT AAATGTGACT CTGATAAAATT CATAGACATT 3480  
 5 TCCACATATG CT 3492

## (2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 1973 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

ATCTAGCGGT ACAAGCGTCT TGGAGGCTAG TATGTTGAAC ATTGTAAACC CTGAAGATCA 60  
 20 CTTCGTTGTC ATTGTTTCAG GTGCCTTTGG TAACCGATTT AACCAAATTG CACAACTTA 120  
 TTACAAAAT GTGCATATTT ATGACGTAAC ATGGGGAGAA GCTGTAGATG TCAAAGATTT 180  
 CATCAATTTT CTTTCAACTT TAAATGTTGA AGTTAAAGCA GTATTAGTC AATATTGCGA 240  
 25 AACATCTACG ACAGTGCTAC ACCCTATTCA CGAGTTAGGA AATGCCATTA ATCAATTTAA 300  
 TAGTAATATT TATTTTGTAG TTGACGGCGT AAGTtGCATT GGTGCTGTTG ATGTTGACAT 360  
 TAACAAAGAT AAAATTGATG TACTTGTTTC TGGTAGTCAA AAAGCAATTA TGTACCTCC 420  
 30 AGGATTAGCT TTTGTAGCTT ATAGCCACCG TGCAAAAGAA CATTTCAAAG AAGTAACTAC 480  
 GCCAAAATTT TATCTAGACT TAAATAAATA CATTTCGTCA CAAGCTGACA ATTCTACACC 540  
 GTTCACACCA AATGTGTCTT TATTTAGAGG TGTAATGCA TACGTTGAAA CCGTAAAAGC 600  
 35 AGAAGGTTTC AATCACGTAA TAGCACGACA CTATGCAATT AGAAATGCAT TAAGAAGCGC 660  
 CTTAAAGCA TTAGATTTAA CTTTATTAGT CAATGATAAA GATGCATCTC CAACGGTTAC 720  
 40 AGCATTCAAA CCTAATACAA ATGATGAAGT GAAAATAATC mAAGATGAAC TTAAAAATnG 780  
 CTTTAAATA ACAATTGcng GTGGTCAAGG CCATCTTAAA GGTCAAATTT TnAGAATTGG 840  
 TCATATGGGG AAAATTAGTC CTTTCGATAT TTTATCGGTA GTATCTGCTT TAGAAATTAT 900  
 45 TTAACTGAA CACCGTAAAG TTAATATAT CCGTAAAGGT ATATCAAAAT ATATGGAGGT 960  
 TATTCATGAA GCAATTTAAT GTACTCGTTG CAGATCCCAT ATCAAAAGAT GGTATCAAAG 1020  
 CATTATTAGA TCACGAACAA TTCAATGTAG ATATTCAAAC TGGCTTGTCC GAAGAAGCAT 1080  
 50 TAATCAAAAT TATACCTTCA TACCATGCTT TAATCGTTCG TAGTCAAAC ACGGTTACTG 1140  
 AAAATATCAT AAATGCTGCT GATTCTTTAA AAGTAATCGC ACGCGCCGGT GTTGGTGTAG 1200

55

5      GTAATACGAT TTCAGCTACT GAACATACAC TGGCAATGTT ATTATCAATG GCACGAAATA      1320  
       TTCCGCAAGC ACACCAATCA CTTACAAATA AAGAATGGAA TCGAAATGCA TTTAAAGGTA      1380  
 10      CTGAGCTTTA TCATAAAACA TTAGGTGTCA TTGGTGCTGG TAGAATTGGT TTAGGTGTTG      1440  
       CTAAACGTGC GCAAAGTTTC GGAATGAAAA TACTAGCTTT TGACCCTTAC TTAACGGATG      1500  
       AAAAAGCAAA ATCTTTAAGC ATTACGAAGG CAACAGTTGA TGAGATTGCC CAACATTCTG      1560  
 15      ATTTGCTTAC ATTACATACA CCACTAACAC CTAACACAAA AGGCTTAATT AATGCTGTCT      1620  
       TTTTTGCCAA AGCAAAACCT AGTTTGCAA TAATCAATGT GGCACGTGGT GGTATTATTG      1680  
       ATGAAAAGGC GCTAATAAAA GCATTAGACG AAGGACAAAT TAGTCGGGCA GCTATCGATG      1740  
 20      TGTTTGAACA TGAACCTGCA ACTGACTCGC CTCTTGTTGC ACATGATAAA ATTATTGTTA      1800  
       CACCTCATTT GGGTGCTTCA ACAGTCGAAG CTCAAGAAAA AGTGGCAATT TCTGTTTCAA      1860  
       ATGAAATCAT CGAAATTTTA ATTGATGGTA CTGTAACGCA TGCAGTGAAT GCACCTAAAA      1920  
       TGGACTTAAG CAATATAGAT GATACTGTAA AATCATTCAT CAATTTAAGC CAA      1973

(2) INFORMATION FOR SEQ ID NO: 36:

25      (i) SEQUENCE CHARACTERISTICS:  
           (A) LENGTH: 7620 base pairs  
           (B) TYPE: nucleic acid  
           (C) STRANDEDNESS: double  
 30           (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

35      GGTGTTTCAG ATGTCACTGG TTGATTTTTA ATTGTAGACG GGTATTTTGG GCTTTTCGCCA      60  
       TATTTATTTG CCGGCTTACT GTCAAAGCAT AGGAATACTA TCATAACAAT TGTTAGGCCT      120  
       AAATGAACAA AATAAGAAG TACTAACAAA ATATTAAGAC CCATCGGCAT TAATGTAAAA      180  
 40      TCACTGTCAT AATAACTATC GATAATCTGT AATACTATAT AAAATATAAT ACTGAATACT      240  
       GTCATAATCA TTGGAAATAA CATTGTTCTT GATATATCGT GAAATCTTCG AACGCACAAC      300  
       GCTAAATTTG GAATAAACGT TGCCAAACTA TAGACAAAAG TATACACAGA TGTAAGGATA      360  
 45      ATCATCAATA TACTCATAAC TATTAATGTT TCGTTATCCG CCGCTATAGA AATAAAGAAT      420  
       AGAAATAGGT TTATTATTAG CACACACACA GCTGGAACCA TAAGTATCAA ATGCCATAGT      480  
       GCCATATACC AATATTCACT ACGTCTTGAT CTCCCCTTAA AATTTACATA ATTTTTCCAA      540  
 50      AATAAAACGA ATGATTTTCAT AAAACCTACT TGAGGTAATT GTTCCATTGT AATCTCCCTT      600  
       TCGTTAATCA TATTTATATT TTTAATTATT GTTACCGTTA TAATTTACAA GATTCATTAT      660

55



EP 0 786 519 A2

	GTAAAAATGAA AACCCGCTAC AAGTACACAT CTATATGGAG ACTCATTTGA AAGTCAACGC	780
	TTCGTTAAC TACTAAAAA TATGTCATAC TGCAATGTTC ACGTTTAAAA GAGTCTCAAT	840
5	CTATGCAAAT AAAATATTCC ATAACAAAGT ATATACTTTA CATTTTATA ATTCTTAACA	900
	ATACTATTTT ATCAAACATT TACCACAATA AAAATATCTT TTTCATTTTT ATTTAAATTA	960
	ATCATATAAT TCGGAGGAGA ATATTATGGA TTTCGTTAAT AATGATACAA GACAAATTGC	1020
10	TAAAAACTTA TTAGGTGTCA AAGTGATTTA TCAGGATACC ACTCAAACGT ATACAGGCTA	1080
	CATCGTGGAA ACGGAAGCTT ACTTAGGTTT GAATGATCGT GCGGCTCATG GCTATGGCGG	1140
15	TAAAATAACA CCTAAAGTCA CGTCATTATA TAAACGTGGT GGTACAAATT ATGCACATGT	1200
	CATGCATACG CATTTACTCA TTAATTTTGT AACAAAATCT GAAGGTATAC CTGAAGGCGT	1260
	ACTTATCCGC GCAATTGAAC CAGAAGAAGG TTTATCCGCT ATGTTCCGTA ACAGAGGTAA	1320
20	GAAAGGCTAC GAGGTAACGA ATGGCCCAGG AAAATGGACT AAGGCATTTA ACATTCCACG	1380
	GGCTATCGAT GCGGCTACGT TAAATGACTG TAGATTGTCT ATTGATACTA AGAATCGTAA	1440
	ATATCCTAAA GATATTATTG CTAGTCCACG AATCGGTATT CCAAATAAAG GTGATTGGAC	1500
25	ACATAAATCT TTACGTTACA CAGTGAAAGG TAATCCATTT GTGTCTCGCA TCGGTAAATC	1560
	AGATTGTATG TTTCCCGAAG ATACTTGGAA ATAAATGCCA TCTTTCATTG ATTACTATCA	1620
	TGAAAAATGAA ATCTATCTCC TTATAAGTCA ATCAATCGTG CCGTCAACAT GCGGATGGGT	1680
30	TGATTGTTTT TCTTGTATC CATCATATTT TTTGATTCAT CTCCTCTTAT TGAAC TTGTT	1740
	CTTAATTATA AAATATAACA ATAGAATTAT TTATAATTAT TAAATTTAGA TGCATTAATA	1800
35	TTATTGATAT TATTTTCAA AACTAGAAAT ATTGATTGTG TGCATGTATA ATGTTAAAAG	1860
	CGCCCTTTTA TAACGCTTAC ATATAAAAGC TTATTTAGGG AGAGGGATAT TCAACAAGGG	1920
	GGATTTGAAA ATGATAGAAC TTAATGCAAT TACAACATTA TGTTTAGCTT GTATCCTTTA	1980
40	TTTACTTGGT AAGGCTATCG TTAATCACGT TAATTTTTTA AAACGTATTT GTATACCAGC	2040
	ACCAAGTATT GCGGCTTAA TCTTTGCTAT TTTAGTTGCG GCTTTGGATT CATTTGGCAT	2100
	GGTTAAGATT AAATTAGATG CTTCAATTCAT TCAAGATTTC TTCATGTTAG CATTCTTTAC	2160
45	GACAATCGGT CTGGGTGCAT CATTGAAATT ATTTAAATTA GGTGGCAAAG TCTTGCTATT	2220
	ATACTTTATG TTTTGTGCTA TCATTTAGT CATTCAAAAC ATAGTTGGTG TATCACTAGC	2280
	AAAAGTATTA AATATTAAAC CTTTGTTAGG ATTAACAGCA GGTTCCATGT CTATGGAAGG	2340
50	CGGTCATGGT AATGCTGCTG CTTATGGTAA GACAATTCAA GATTTAGGTA TTGATTCCGC	2400
	ACTGACAGCG GCTCTTGCA CTGCAACTTT AGGTCTTGTA TTTGGAGGGC TTATCGGTGG	2460
55		

	ATTTAAAGAT TATAGCCAAG TAGCATATAA CGAACATTTA CATAGTAAAT TTAATGCCAC	2580
5	TGAAGTATTC TTCATTCAAT TTACAATCGT TGTATTCTGT ATGGCAGTTG GAAGTTATTT	2640
	CAGTCATTTG TTTACAGCTC AAACAGGGAT TAATGTTCCA ATTTACGTTG GCTCATTATT	2700
	TGTAGCTGTT ATTGTCCGAA ATATCTCTGA AAGTTTTAAT TTAAATATTG TAGATTTAAA	2760
10	AATTACTAAT CAAATTGGCG ATGTCGCATT AGGTATTTTC TTATCTCTTG CGCTAATGAG	2820
	CATTCAATTA ATCGAAATTT ATAACTTGC TATACCTCTT ATTATTATCG TTTTAGTTCA	2880
	AGTTGTCGTT ATGATTTTAT TTGCTGTTTT AATTTTATTT AGAGGTTTAG GAAAAGATTA	2940
15	TGATGCTGCA GTAATGGTAG GTGGTTTTAT CGGTCATGGG CTTGGTGCAc GCCAAATGCC	3000
	ATGGCAAATT TAGATGTTAT TACTAAAAAA TATGGAACT CACCTAAAGC ATATTTAGTT	3060
	GTACCTATTG TTGGTGCATT CTTAATCGAT TTAATTGGTG TTATAGTCAT TATGGGATTC	3120
20	ATACAATGGT TTAGTTAAAC ACCAACTCA TAAATAAAAG AGGAGGCCTT CGCCTCcTcT	3180
	TTTATTTATC CTCGATGTAT ATTCAAGTTA CGTTGTTCTA TCCATGACAA TATTTCCGGA	3240
	CTAAATACGA TTTGTTTTTG TGTTAAGTCG TCAATATTTT TAGCATCTAA CATCGTCATT	3300
25	ATTGATTTCA TGTGTTCAAT AAATGATTCT ACATAAGCTA CTGTATGTGC AATGCCATTA	3360
	TTTTCAACTT GATTTAAAAA CGGACGTGAC ATACCAGTTG CCTTTGCACC AAGTGCTAAA	3420
30	CTTTAATTG CATCGAGTGG TGTACGTAAA CCACCACTCG CGAAAACIGA AATTCGCTT	3480
	TGATAAGCCG TTGTTTCAAG TAATGACTCA ACTGTAGACT GTCCCATGA TGATAAGTAA	3540
	TCCATATCTT TATTGACAG ACGTTCATTT TCAATATCTA CAAAGTTAGT ACCACCTTTG	3600
35	CCACTAACAT CGACATACTT GACGCCTATT TGTTGTAAGT CATGCATTAA TTCTTTGCTC	3660
	ATACCAAATC CAACTTCTTT TATAATGACT GGAACAGACA CTCGTGATAC AATCGACGCT	3720
	ATATTATCTA ACCAAGTCAC AAATTCACGA TTCCCTTCAG GCATAACTAA TTCTTGAGGA	3780
40	GAATTAACAT GGATTTGTAA CGCTTGCGCC TCAAGTAATT CAACTGCTTC CAAAGCCTTT	3840
	TCTACTGGTA CGTCCGCACC AACATTGCTA AAAATCATGC CTTCAGGATT CATTTTTCGC	3900
	GCAATCGTAA ACGTCTCAGC CATGCGTGGA TTTCTCAATG CCGCATGTGT TGATCCAACT	3960
45	GCCATCGCTA AGCCAGTTTC TCTTGCAACT ACAGCTAGCT TTTCAATTGAT GTTTTTCGTC	4020
	CACTCGCTAC CACCCGTCAT TGCATTAATA TAAACCGGAT ATGCCATCGT TAAGTCAGGC	4080
50	GTCTGTGATG TCAAATCGAT ATCATTTACA TTAATTGATG GGATAGAATG ATGCACAAAA	4140
	CGCATCTTAT CAAAATCTGA ATGCATTGCG TCAGATTGGG CCATTGCTAT TTCAACATGT	4200
55	TCATTTTTTC TCTGTTCTCT TTGAAAATCA CTCATGATTA AACCTACCTT TTCGTCATTT	4260

	ATTACAGCTA AGCAAATATA ATATCCATAA TGTAATGTGA ATGCCGGCAT ATTTACAAAG	4380
5	TTCATACCAT AAATCCCAGC TATGAATGTT AACGGTGAAA ATATAACTGA TACTAATGTC	4440
	AGTACTTGCA TAATACTATT CATTCTAAAT GACGTGTATG ACTCAAAATT TTCTCGTATT	4500
	TCGTTTGTC TTTCTTGAGC AGTACGAATG ATATTACGTT GCTTAATCAA GTGGTCATCG	4560
10	ATATGTTGAA TGTATAGCGA ATGTTTATTA TCTATAATCA AATCACCATT TTGTTTCATT	4620
	GTATCAATTA GCTCTTGCAT AGGAAACAGT ACACGTTTTA CTTTAATCAA ATCCGAACGT	4680
	AACTTAAAGA CACTATCCAT GACCATTTTA TTAAAGCGAT CATCTACATG GCGGTCTTCA	4740
15	AAATGATAAA CACTATCTTC AAGTGCATAT ACAAAGTTGA AATATTTATC AACCATCATA	4800
	TCTAAATTA ATATGACGAC ATCTGCACAA TCTAATTCTG CATCTAATGT ATTCATATAC	4860
	TTATAGACTA CTTTATTTAA TGATTCCAAC GTTTGATGAT GATATGTTAC TAATACATTG	4920
20	TCTTGTATAA AAATATTTAG TGCTATTGGT GAATAGTTTG ACCCCATAAT ACTATGGAAT	4980
	ACTAAGTATT GATAATCTTT ATAAGATTTA TATTTAGCTC GTGGCATACC GTTAATTGCA	5040
	TCATCCACTT CTAATCATT AAAATTAATA TGTCCTTTAA ACCATTCAAT TTCTTGTTCA	5100
25	TTCGGTTTCAT CAAATCATA CCAAACAATA GTCGCATCTT TTGGTATCTC TTTGATATCA	5160
	TCAACTACTT TAAACGGTTC ATATGTAGTT TGATACCGTA TCTTTAAAGC CATCGATACT	5220
30	CCCCCTAAAT AACGAATTCT CTATTATTTT ATCATGAATT AAATAACGTG TATGTCTTAA	5280
	TTTATTTTAG TATGATAGTC ACTAAGGAGA TGGTTATTAT CAAACAACCT TTTACACATA	5340
	CTCAAACCGT AACATCTGAA TTCATTGACC ATAACAATCA TATGCATGAT GCAAATTATA	5400
35	ATATCATTTT TAGTGACGTC GTGAATCGTT TTAATTACAG CCACGGTCTT TCTTTAAAAG	5460
	AACGCGAAAA TTTAGCATAT ACGCTATTTA CACTAGAAGA ACATACGACA TACCTCTCAG	5520
	AATTGTCTCT TGGCGATGTA TTTACTGTTA CTTTATATAT TTATGATTAC GATTATAAGC	5580
40	GGTTGCATTT ATTTTAAACA TTAATAAAG AAGATGGTAC ACTAGCATCA ACAAATGAAG	5640
	TAATGATGAT GGGAAATTAAT CAGCACACAC GTCGTTCTGA TGCTTTTCCT GAATCATTTT	5700
	CAACACAAAT AGCACACTAT TATAAAAATC AATCAACTAT CACTTGGCCT GAACAATTAG	5760
45	GACATAAAAT AGCAATTCCA CACAAAGGAG CATTAATAATG ACAGATGCAT TACAACAAAA	5820
	GATTCATATC GAATTACTAG ATTTATTAGA TGATGTTAAG TTTGAATTAA CAGAATTAAA	5880
50	TGCACAAAAA GGGTTATACA TTAACGGACC AGCAAATCAG CTACTTAAGC GTGGCGTGCA	5940
	TATGGCTTAT GTTCAAGGAC AAAAGCAAGC CATCGATAAT ATTATGACTA TTGTGGAACA	6000
	ACAGCTTGAA AGATCAACAT TTCCTAGAAC ATTATGATAA ATTTCAAAAT GAGGTTGCTC	6060
55		

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45

ATAATTTTTT AGATCAATTT TATCAAATTA AAGGGCAATA CTTTATCATC ACACATATCA 6180  
 ATACACTTAT TGGTGATTTT CACTCAGAAG CTCATTAACA ATTAGTCTAT ATAACCCTTG 6240  
 CTATATTTTC AAAAACAAAA CCCAATTACG TTTTCATGTC AAATATCATC TTGCATGAAA 6300  
 TCGTAACTGG GTCATTTATA TGTTATTAGT TATTTTGTGT TACATCCTCA TCTATCGATT 6360  
 TGGCAATTTG TTTAATAGCT TTATGTGATT GTCTAATTGG ATAAATTGGA AAATCATGTA 6420  
 CCATCTTAGG ATAATCATAA AACTCAATGT ATTGATGATG TTGCAACATC ATTTGTTCAA 6480  
 ATAGCTTCAT ATCAGGATGT GTCATTTTAC GTCCACCACC AAACATATAA ACTGGTGGCA 6540  
 ATCCTTCTAT TGTGCCATTA ATTGGCGATA TGCGCTTATC TGTTAATGGT AGGCCATTCTG 6600  
 CCCATTTTTT CATAATCTCA TTGACACCAA ACTGACTTAG aACCGCATCT TGTTGATTA 6660  
 AGGCGTCCGA AATATCTTTA TTAGATAGTG TTGCATCTAA AATTGGTGAG ATTAAATACA 6720  
 ATTTATTCGG TAATGGCTGT TGATTakCTA AAAGAGATTG TACAAAGGAT AATGCCAGTG 6780  
 CACCACCTGA ACCATCACCC ATGACTACGA CATTTTGATG TCCTACTTCA GATACTAATT 6840  
 GaTCATAAAC ACGTTGTATC GCTTGGnAAA GTATCGTCaA TATGnAAACT CTGGTGTCTT 6900  
 TGGATAGATA GGCAGTACAA CCTCATATAA TGtACTTAAA GTGATTTTAT CCCAACAATC 6960  
 TCCAATGGAA CGGTGATGGT TGTAAGTGCAT TGAATCCACC GTGAATATAT AAAATTTTCT 7020  
 TATCAATTTG ATGTCTGAAA TTAAAGCGAA AGACTTGCAT ATCATCTAAT GACAATTTTT 7080  
 CTAAATTTGC TTTAACATTT AATGTTGAAG GCTGCTTATG TTTTTTCTA TTTTCAATTT 7140  
 CTCTTTTATA AAAAATCTT TCAACATCTT GATCATTTTT AAACATAATC GAGCGATTGT 7200  
 GAAGCAAATA TTTATTGACA ACGCTATTCA TAACACGGTT TCTAATCAAT GTCTTAACCT 7260  
 ACCTTTATAT ATTTTATGTA TCCAATGATk GTCTATCCCC TACATTCTTT GCCAAAAAAA 7320  
 GTATATAATG TAGAAGATAT TTTCTTTTTC ACTTTCAAAT TTAAGACTAC AATTGAACAG 7380  
 TGATTTTTCA TCATTATAAC AGACAAC TAG ACATATTGAT AAGTAAAGAA AAGAACTTTA 7440  
 TACGGAGGTA CCTTGCATGA CAAATCCAAA TCAACGATTA GAACCATTG ATGAGACATT 7500  
 TCAACAACCG AATATTCATC GTGGTAAGCG ATATGGTAAG AAAAAACGTT CATTGGTAAG 7560  
 CATGATTATT CAAATCATTG TTGTWATATT AACCACCATC GCTGGAATAC AGCATGGTGG 7620

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9834 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

	GTCATtACCG amTTTtTAG AaTCATTtTAA AGATGATAAA TATACAAACG TTGGTAATTT	60
5	AAAAGAAGTG AATTTTGATA AAATTGCTGC GACGAAACCC GAAGTAATCT TTATCTCTGG	120
	ACGTACAGCT AATCAAAAGA ATTTAGATGA ATTCAAAAAA GCTGCACCTA AAGCGAAAAT	180
	TGTTTATGTT GGTGCAGATG AAAAGAACTT AATTGGTTCA ATGAAACAAA AACTGAAAA	240
10	TATCGGAAAA ATTTACGATA AAGAAGATAA AGCTAAAGAA TTAAATAAAG ATTTAGATAA	300
	CAAAATTGCT TCAATGAAAG ATAAAACGAA AAACCTTCAAT AAAACTGTTA TGTATTTACT	360
15	AGTTAACGAA GGTGAATTAT CAACATTGCG ACCTAAAGGT CGTTTTGGTG GATTAGTTTA	420
	CGATACATTA GGATTCAATG CAGTTGATAA AAAAGTAAGT AATAGCAATC ATGGACAAAA	480
	TGTTTCTAAC GAATATGTTA ATAAAGAAAA TCCAGATGTT ATTTTAGCGA TGGATAGAGG	540
20	TCAAGCGATA AGTGGTAAAT CAACTGCGAA ACAAGCATT AATAATCCTG TATTAAAAAA	600
	TGTTAAAGCA ATTAAGAAG ACAAGTATA TAATTTAGAT CCTAAATTAT GGTACTTTGC	660
	AGCTGGATCA ACTACAATA CAATTAAACA AATTGAGGAA CTTGATAAAG TTGTAAAATA	720
25	ATTTTAAAAG AGGGAACAA TGGTTAAAGG TCTTAATCAT TGCTCCCCTC TTTCTTTAA	780
	AAAAGGAAAT CTGGGACGTC AATCAATGTC CTAGACTCTA AAATGTTCTG TTGTCAGTCG	840
	TTGGTTGAAT GAACATGTAC TTGTAACAAG TTCATTTCOA TACTAGTGGG CTCCAAACAT	900
30	AGAGAAATTT GATTTTCAAT TTCTACTGAC AATGCAAGTT GGCGGGGCCC AAACATAGAG	960
	AATTTCAAAA AGGAATTCTA CAGAAGTGGT GCTTTATCAT GTCTGACCCA CTCCCTATAA	1020
35	TGTTTTGACT ATGTTGTTTA AATTTCAAAA TAAATATGAT AGTGATATTT ACAGCGATTG	1080
	TTAAACCGAG ATTGGCAATT TGGACAACGC TCTACCATCA TATATTCATT GATTGTAAAT	1140
	TCGTGTTTGC ATACACCGCA TAAGATTGCT TTTTCGTTAA ATGAAGGCTC AGACCAACGC	1200
40	TTAATGGCGT GCTTTTCAAA CTCATTATGG CACTTATAGC ATGGATAGTA TTTATTACAA	1260
	CATTTAAATT TAATAGCAAT AATATCTTCT TCGGTAAAAT AATGGCGACA scgTGTTTCA	1320
	GTATCGATTA ATGAACCATA AACTTTAGGC ATAGACAAAG CTCCTTAACT TACGATTCTT	1380
45	TTGGATGTTT ACCAATAATG CGAACTTCAC GATTTAATTC AATGCCAAAT TTTCTTTGA	1440
	CGGTCTTTTG TACATAATGA ATAAGGTTTT CATAATCTGT AGCAGTTCCA TTGTCTACAT	1500
	TTACCATAAA ACCAGCGTGT TTGGTTGAAA CTTCAACGCC GCCAATACGG TGACCTTGCA	1560
50	AATTAGAATC TTGTATCAAT TTACCTGCAA AATGACCAGG CGGTCTTTGG AATACACTAC	1620
	CACATGAAGG ATACTCTAAA GGTTGTTTAG ATTCTCTACG TTCTGTAAAA TCATCCATTT	1680

55

	AGTGTTCTTT TTGAATAATG CTATTACGAT AATCTAACTC TAATTCTTTT GTTGTAAGTT	1800
	TAATTAACGA GCCTTGTTTCG TTTACGCAAA GCGCATAGTC TATACAATCT TTAACCTCGC	1860
5	CACCATAAGC GCCAGCATTG ATATACACTG CACCACCAAT TGAACCTGGA ATACCACATG	1920
	CAAATTCAAG GCCAGTAAGT GCGTAATCAC GAGCAACACG TGAGACATCA ATAATTGCAG	1980
	CGCCGCTACC GGCTATTATC GCATCATCAG ATACTTCGAT ATGATCTAGT GATAATAAAC	2040
10	TAATTACAAT ACCGCGAATA CCACCTTCAC GGATAATAAT ATTTGAGCCA TTTCCTAAAT	2100
	ATGTAACAGG AATCTCATTT TGaTAGGCAT ATTTAACAAAC TGCTTGTAAT TCCTCATTTT	2160
	TAGTAGGGGT AATGTAAAAG TCGGCATTAC CACCTGTTTT AGTATAAGTG TATCGTTTTA	2220
15	AAGGTTCAAT AACTTTAATT TTTTCATTTG GGATAAGTTG TTGTAAAGCT TGATAGATGT	2280
	CTTTATTTAT CACTTCTCAG TACATCCTTT CTCATGTCTT TAATATCATA TAGTATTATA	2340
20	CCAATTTTAA AATTCATTTG CGAAAATTGA AAAGAAAGTA TTAGAATTAG TATAATTATA	2400
	AAATACGGCA TTATTGTCGT TATAAGTATT TTTTACATAG TTTTCAAAG TATTGTTGCT	2460
	TTTGCATCTC ATATTGTCTA ATGTGTAAGC TATGTTGCAA TATTTGGTGT TTTTTGTAT	2520
25	TGAATTGCAA AGCAATATCA TCATTAGTTG ATAAGAGGTA ATCAAGTGCA AGATAAGATT	2580
	CAAATGTTTG GGTATTCATT TGAATGATAT GTAGACGCAC CTGTTGTTTT AGTTCATGAA	2640
	AATTGTAAAA CTTCGCCATC ATAACTTTCT TAGTATATTT ATGATGCAAA CGATAAAACC	2700
30	CTACATAATT TAAGCGTTTT TCATCTAAGG ATGTAATATC ATGCAAAATT TCTACACCTA	2760
	CTAAAATATC TAAAATTGGC TCTGTTGAAT ATTTAAAATG aTGctACCGC CAATATGTTT	2820
	TGTATATTTT ACTGGGCTGT CTAAGAGGTT GAATAATAAT GATTCAATTT CAGTGTATTG	2880
35	TGATTGAAAA CAATTAGTTA AATCACTATT AATGAATGGT TGAACATTTG AATACATGAT	2940
	AAAGTcCTTT GATATTGAAA ATTAATTTAA TCACGATAAA GTCTGGAATA CTATAACATA	3000
40	ATTCATTTTC ATAATAAACA TGTTTTTGTA TAATGAATCT GTTAAGGAGT GCAATCATGA	3060
	AAAAAATTGT TATTATCGCT GTTTTAGCGA TTTTATTTGT AGTAATAAGT GCTTGTGGTA	3120
	ATAAGAAAA AGAGGCACAA CATCAATTTA CTAAGCAATT TAAAGATGTT GAGCAAAAAAC	3180
45	AAAAAGAATT ACAACATGTC ATGGATAATA TACATTTGAA AGAAATTGAT CATCTAAGTA	3240
	AAACTGATAC AACTGATAAA AATAGTAAAG AATTTAAGGC ACTACAAGAA GATGTTAAAA	3300
	ACCATCTCAT ACCTAAATTT GAAGCATATT ATAAGTCAGC AAAAAATTG CCTGATGATA	3360
50	CAATGAAAGT TAAGAAATTA AAAAAAGAAT ATATGACGCT TGCAAATGAG AAGAAGGATG	3420
	CGATATATCA ATTAAAAAAA TTCATAGGTT TATGTAATCA ATCTATCAAG TATAACGAAG	3480

55

	AATTAGCTGA TAATAAAAGT GAAGCAACTA ATCTTACGAC AAAATTAGAA CATAATAATA	3600
	AAGCGTTAAG AGATACTGCG AAGAAGAACC TAGATGATAG TAAAGAAAAT GAAGTAAAAG	3660
5	GCGCGATTAA AAATCACATT ATGCCAATGA TTGAAAAGCA AATTACCGAT ATTAACCAAA	3720
	CTAATATTAG TGATAAGCAT GTTAATAATG CAAGGAAAAA CGCAATAGAA ATGTATTACA	3780
	GTCTGCAGAA CTATTATAAT ACACGTATTG AAACAATAAA GGTTAGTGAG AAGTTATCAm	3840
10	AAGTCGATGT AGATAAGTTG CCGAAAAAGG GTATAGATAT AACTCACGGC GATAAAGCCT	3900
	TTGAAAAAAA GCTTGAAAAA TTAGAAGAAA AATAACTATA ATCATTTTTTC AAAGTTAAAA	3960
15	ATTTTGAATT TATGGTTAAC ATGTCAACTT ACTATGTGTA TAATGGTAAA CATTGATATT	4020
	AACTATATGT ATAAAAATGT CACGCAGATG CTATTTAAAT GTGATAAATA TTTTGTAGAGG	4080
	TGAATAGAGT GGCTATAAAG CTAAGTTCAA TTGACCAATT TGAACAGGTT ATTGAGGAAA	4140
20	ATAAATATGT TTTTGTATTA AAACATAGTG AAACCTGTCC AATATCGGCA AATGCGTACG	4200
	ATCAATTTAA TAAATTTTTA TATGAACGCG ATATGGACGG TTATTATTTG ATTGTCCAAC	4260
	AAGAACGCGA TTTGTCAGAT TATATTGCTA AAAAAACGAA CGTTAAACAT GAATCACCTC	4320
25	AAGCATTTTA TTTTGTAAAT GGTGAAATGG TTTGGAATCG AGACCACGGT GATATCAATG	4380
	TGTCGTCATT AGCACAAGCA GAAGAATAAT GAACTATAG GGTGGAACA TTTTGCCTTA	4440
	CACTACTAGA CGTGAATAGC ACAACTTAAA TTCGTGTGAA TCAGAGTAGT TTGGCTATAA	4500
30	TGATGTTCTG ACCTTTTATT TTATGTCACC TTTAGAAGCA GTTAAGTTAG TACTTTTTTTA	4560
	CAAACATATG TATAATATAT TCGAGTATTT TTATTGAAAa tATTTTGGAA AACGACGAAT	4620
35	CCAATAAGAA AATTTAAACA TGATTGTGTA GTTAGTTTAA TAGGAAATAT ATGCTAAACC	4680
	AAAAGAAGCA TATTGTTATT TACTGGAATA ATTAATAATC ATGTCATGTT AAATGTTAGC	4740
	ATATAATCAC GAGATAAAAT CTAAAATTTA AGATTAATCT TTTATGAATA AAAAACGTAT	4800
40	CACAACAAAT AATAAAGTAA GGTGGTCAAG GTTATGAAAG TATTAGTAGC CATGGATGAG	4860
	TTTCATGGAA TTATTTCAAG TTATCAAGCT AATAGATATG TTGAAGAGGC AGTTGCAAGC	4920
	CAAATTGAAA CTGCAGATGT AGTTCAAGTA CCATTGTTTA ATGGAAGACA TGAATTATTA	4980
45	GATTCTGTAT TTTTATGGcm ATCTGGGcaA AAGTATCGTA TACCAGTACA TGATGCAGAT	5040
	ATGAATGAAG TTGAAGGTGT TTACGGACAA ACTGATACAG GGATGACCGT TATCGAGGGG	5100
	AATTTATTTT TAAAGGTAA AAAACCAATT GTTGAACGAA CAAGTTATGG TTTAGGAGAA	5160
50	ATGATTAAAC ATGCATTAGA TAACGACGCA AAACATGTTG TAATTTCACT AGGTGGGATT	5220
	GATAGTTTTG ATGCTGGTGC AGGTATGTTA CAAGCATTAG GTGCTCAATT CTATGATGAC	5280
55		

	GATATGTCGA	ACTTACACCC	TAAAATGGAA	ACAGCAAGAA	TTCAAGTAAT	GTCGGATTTT	5400
	TCAAGTCGAT	TATATGGTAA	GCAAAGTGAA	ATCATGCAAA	CTTATGATGC	GCATCAGTTG	5460
5	AATCATAATC	AAGCAGCAGA	AATCGATAAT	TTAATTTGGT	ATTTTAGTGA	GTTATTTAAA	5520
	AGTGAATTGA	AAATTGCAAT	TGGTCCAGTT	GAACGTGGTG	GTGCTGGTGG	TGGAATTGCA	5580
	GCAGTCTTGA	ATGGACTGTA	TCAAGCTGAA	ATATTAACCA	GTCATGCATT	AGTAGACCAA	5640
10	CTAACACATT	TAGAAAATTT	AGTTGAACAA	GCGGATTTAA	TTATTTTGG	AGAAGGATTA	5700
	AATGAAAATG	ATCAGTTGCT	AGAAACGACA	ACATTGCGTA	TTGCAGAACT	TTGTCATAAA	5760
15	CATCAAAAGG	TTGCCATTGC	AATTTGTGCA	ACTGCTGAAA	AGTTTGATTT	ATTTGAATCA	5820
	CAAGGGGTTA	CAGCAATGTT	TAATACATTT	ATCGATATGC	CAGAAACTTA	TACTGACTTT	5880
	AAAATGGGtT	ACAAATTAGG	CATTATACGG	TTCAGTCTTT	AAAACACTTG	AAAACACATT	5940
20	TTAATGTTGA	GGTTTAGTAA	AGAAGGACTA	AATTGGTGAT	GCTGTCATGA	TGGTTAATAA	6000
	CATTATGAT	GGTTAGCAAA	ACGAATTAGA	AGATCGAAAG	TATACGTAAA	AAATATGAAA	6060
	AATCACGCTA	TCATTGCACT	GAATGTTAGC	GTGATTTTTA	TATATTAATT	AAGCCTGAGT	6120
25	TGAAGTAGTA	TATAATCGTT	GGTTTTTAGT	GATTTTCAGC	GATATCTTCT	ACAATTCCAA	6180
	TGATTACTTG	TACTGCTTTT	TCCaTAACAT	CAATGGATGC	aTATTCATAT	GGGCCGTGGA	6240
	AGTTACCGCA	ACCTGTAAAG	ATGTTTGGAG	TTGGTAACCC	CATAAATGAC	AATTGTGAAC	6300
30	CATCTGTACC	ACCGCGAATA	GGTTCAGTGT	TTGCTGGAAT	ATCTAATTTG	GCAAAGACAC	6360
	GTTTAGGTAT	ATCAATAATA	TGAGGCAATG	GTAATATTTT	TTCTGCCATA	TTGAAATATT	6420
	GATCCGATAT	ATCAACTTTA	ACTGGATAAT	TTTCAAAATG	GGCATTGATA	TCGTCACGTA	6480
35	TTTCTAAAAT	ACGTTTCTTA	CGCAATTCTGA	ATTGTTTTTT	ATCATGATCA	CGAATAATGT	6540
	ATTGCAAAGT	TGCTTTTTCA	ACAGTTCCTT	CAAAGTTCAT	TAAGTGATAA	AAGCCTTCGT	6600
40	ATCCTTCTGT	TCGCTCCGGA	ACTTCACTAT	CAGGTAGCAA	ACTATCGAAT	TGTTACCTA	6660
	AACGTATTGC	GTTTACCATT	GCATTTTTAG	CTGAACCAGG	ATGAACATTT	ACACCGTGGC	6720
	ATGTAATAAC	CGCTTCAGCA	GCGTTAAAGC	TTTCATATTG	TAATCTCCA	TATTGACTAC	6780
45	CATCCATAGT	ATAAGCAAAA	TCAGCATTGA	AGCGGTCAAC	ATCAAATTTA	TGTGGACCAC	6840
	GACCGATTTC	TTCGTCTGGT	GTAAATCCAA	TGCGAATGGT	ACCATGTTTA	ATTTCTGGAT	6900
	GTTCTTGTA	ATAACAAATA	GCTTCCATAA	TTTCCACAAT	ACCGCTTTA	TCGTCTGCAC	6960
50	CTAGTAACGA	TGTACCATCA	GTTACCATTA	ATGTATGACC	AACTAAACTG	TTAAGTTCTG	7020
	GAAATACTTT	AGGATCTAAG	ACACGTTTAG	TATTGCCTAG	TTTGTATGGC	TTACCATCAT	7080

55



	GCGCCAAAAA	TCCAACGTGT	GGGACGTCGA	CATCGATGTT	ACTTTCTAAT	GTAGCAAATA	7200
	AGTAGCCATT	TTCATCTAAA	TCAGTTGGCA	ATCCTAATTG	TTGTAATTCT	TTTTCTAATA	7260
5	AATGTAACAA	ATCCCATTCG	TTTTCAGTTG	AAGGTGTTGT	TGTAGATTTT	GGATCAGATT	7320
	GCGTATCAAT	TGTCGTATAT	CTTGTTAATC	TATCTATCAA	TTGGTTCTTC	ATTATATTCG	7380
	ACCCCTTAAA	CTCTATTATT	CATGTTGTAA	GATTTTATAT	ATGTCTTACC	TTTGATTTTA	7440
10	CCATACAGTT	GTTTGATACG	TGTGTATAGG	TAATATAGAA	TTTCAGAAAC	TAATATACCG	7500
	AAAGCAATCG	CACCTGAAAT	CAGTGTACTT	CTAAAAATGT	ATTTACAGCA	CTTGTATAAT	7560
15	CATTTGATAC	TAAAAAACGA	GTCGCTTGAT	AAGCTGCACC	ACCAGGTACT	AATGGTATAA	7620
	TGCCTGGCAC	TATGAATATA	ATTACCGGTC	GTTTATATCT	GCGACTCATA	GTATGACTCA	7680
	TTAAGCCTAA	AATTAAGCTT	CCCAAAAATG	AAGCGCCAAC	TTTTCCAAAC	TCTAAATCTA	7740
20	CCGTTAATTG	GTAAATCGTC	CATGCAATGG	CACCCACAAA	TCCACATGCT	ACTAAGAGGC	7800
	GTTTGGGTGC	ATTGAAAATG	ATAGAGAAAA	GTACTGTTGA	TATAAAGCTG	ATTGTAAAAT	7860
	GAAATAAATA	AAATAGCATG	CTTTAACAGT	CCTTCCTTAA	ATGATTAATA	AAACGATTGC	7920
25	GACACCAGCA	CCGATTGCGA	ATGCTGTTAA	TGCAGCTTCA	ACACCGCGAG	ACATACCTGC	7980
	AAGTAATTCA	CCCGCTAATA	AATCTCGAAT	GGCATTGGTA	ATTAATATAC	CAGGGACAAG	8040
	TGGCATGACA	CTGGCTATAG	TAATGATATC	TTGATTGGTT	GCAATGCCTA	ATTTAGTAAA	8100
30	TGTGGCTGCA	ATGGATATGA	CCACAGCGGC	TGCAACAAAC	TCTGAGAAAA	ATTTAATTTG	8160
	TATATAGCGT	TGCACAAAGC	TGAATGTTAA	AAATGCGGAT	CCGCCAGCAA	TGACTGCAAT	8220
	CCAACAATCT	GATGCGACAC	CACCAAACAT	AAATAGGAAG	AAGCCACATG	CAATGGCAGC	8280
35	TGCAAAAGAA	TTCGTTAAAA	AAGAATATTG	TAATGATGCA	TGCTGTAAAT	GAATAAATTC	8340
	AGATTTAGCT	TCATCAATTG	TGAGTTCTTT	ATTTGATATT	TTACGTGAAA	GACTATTTCGT	8400
	TAAAGCGATT	TTCTCTAAAT	CTGTTGTACG	CTCTTGATCA	CGAATTAATC	TTGTACTTGT	8460
40	TCGATCGTTT	AATGAAAAAA	TAATTGCAGT	TGAACTGACA	AACTATATG	TATTATGAAG	8520
	ACCATAACTA	TGTGCGATAC	GGTTCATTGT	ATCTTCAACT	CGATATGTTT	CAGCACCTGA	8580
45	TTCaAGTAAA	ATTCTACCTG	CAATTAATAC	AACATCAATC	ACTTTGTTTT	CATCTATAAT	8640
	TGTGATTGAA	TCTGGCATAT	CAATTCACCT	CCAATGATAT	GTGTTATTTA	TTGAACAAT	8700
	TGaAGTTTAC	AACTTGTTGT	TACAACTTTC	AATAGTGAGA	CTTGTGTGTA	GTATGATGAA	8760
50	CTTGTATGGT	TCAAATTTAA	ATAAGAAAAA	CTGTTAATCT	TTGCTATTAT	ACTATGATTT	8820
	AATAATAGCA	AAGGATTAAC	AGTTTTGTCT	TTGTTATAAA	TTGATAATAG	GGTTAAACAT	8880
55							

TTTACGCTGT GATTTTGGAT CGTCATCTGT TAAATAACCA ACACCGATAG ACACTGACAA 9000  
 TTTAATAACT TCTTTGTTTG GTAAATGGAA TGATGATTTT TCAACACCCG AACGAATATT 9060  
 5 TTCAGCTAAT TTAACACTTT GATCAAGTGA ATAATTGTGA ATGACAACTG AGAACTCTTC 9120  
 GCCACCATT CTAAAAATTT TAAATTGATT CGGCACATAG TTTTAAAGTA ATTGAGACAT 9180  
 TTGTTTTAAT ACAGCATCAC CTGATTTGTG TGAGTAGGTA TCATTGaCAT CTTTAAATCC 9240  
 10 ATCGATATCG ATTAATAATA ATGCGATACT TTGATGTTCT TTTTCAGCTT TTCGTGAAAT 9300  
 TTCATTTAAA TGTCTATCAA ATTCTTTTAC ATTACCTAAG CCTGTTAAGT AATCATATTT 9360  
 15 ATCTTCGTTT TCATAACGAT TTACGAGTGA GAAGAAATGC CAAATATCGA CAAATGTTAT 9420  
 CGCTGAAGCT AAAGTGATAA TTAATGAAAT TGGTATTAAA ATGATAACTT CCGATAGTGT 9480  
 GTAAATAGGA CTCACTAACG CGACACCAAA TAAAATGATT ATTGTAACAA CATTAAGTAT 9540  
 20 TAATAATGAT AGCACATCAT TTTGTTTTAA AAATGGTCCA ATAGCACTTG TTAGTGCAGC 9600  
 AATAACAATC AACGTAACAC CGTACATAAT CGAGTTGTTA AATACTACAA TTTCAACAAT 9660  
 TGCTACAATT ACTGTGGCAG ATAATGTATA GACCATATTT GTAAATCTAC CTAAAAACAA 9720  
 25 TAAAGGAACG AATGTTAAGT GAATTAAATA ATCTTCACGA TAAGGGATAG GGTAGACAGA 9780  
 TAATAATAAT GATACGATTG TCATTAAAC AGTGACATAA GCCTTAGAAA AAAC 9834

(2) INFORMATION FOR SEQ ID NO: 38:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 23439 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

40 TCTCAATCAG ATGAAAAATT GCATATCGTA GGTTTTACAG AAAGTGCAA ATATAATGCG 60  
 TCATCAGTCA TTTTCACGAA TGACGCTACC ATTGCCAAGA TCAATCCTAG ATTGACTGGA 120  
 GATAAAATTA ATGCAGTTGT TGTACGTGAT ACAAATTGGA AAGACAAAA ATTAAACCAA 180  
 45 GAGCTTGAAG CGGTAAGTAT TAATGACTTT ATTGAAAATT TACCAGGTTA TAAACCACAG 240  
 AACTTAACAT TAACTTTAT GATTTCAATC TTATTTGTCA TTTGAGCTAC AGTTATAGGC 300  
 ATTTTCCTAT ATGTCATGAC ATTACAAAAG ACGAGTTTAT TTGGCATATT AAAAGCTCAA 360  
 50 GGATTTACGA ATGGCTATTT GGCGAATGTG GTAATTTTCG AGACGGTCAT ATTAGCACTA 420  
 TTTGGTACGG CATTTGGCTT ACTGTTAACA GCGGTTACAG GTGCATTTTT ACCTGATGCA 480

55

	TCTGTATTAG GAAGTTTATT CTCCATTTTA ACAATTAGAA AAATAGATCC GTTAAAGGCG	600
	ATTGGGTAGG AGGTGTAGCA AATGTTGAAA TTTGAAAATG TAACAAAGTC ATTTAAAGAT	660
5	GGGAATCGTA ACATTGAAGC GGTAAAGAT ACAAATTTTG AGATAAATAA AGGTGATATT	720
	ATAGCATTGG TTGGACCTTC TGGCTCTGGT AAAAGTACAT TTCTAACTAT GGCAGGTGCT	780
	TTACAAACAC CGACATCTGG GCACATTTTA ATCAATAACC AAGATATTAC GACAATGAAG	840
10	CAAAAAGCAT TGGCAAAAGT TAGAATGTCT GAAATAGGTT TTATTTTACA AGCTACAAAC	900
	CTTGTACCAT TTTTAACGGT AAAGCAACAA TTTACATTAT TGA AAAAGAA AAATAAGAAT	960
15	GTTATGTCTA ATGAAGACTA TCAGCAACTT ATGTCACAAT TAGGTCTAAC TTCATTGCTT	1020
	AATAAGTTAC CTTCAGAAAT TTCAGGTGGT CAGAAACAAC GTGTGGCGAT AgCaAAGCGT	1080
	TATATACGAA TCCGTCGATT ATTTTAGCGG ATGAACCTAC CGCGGCGTTA GATACTGAAA	1140
20	ATGCGATTGA AGTCATTAAA ATTCTACGTG ATCAAGCCAA ACAAAGAAAG AAAGCATGTA	1200
	TTATTGTTAC ACATGATGAA CGACTTAAAG CATATTGTGA TCGTTCATAT CATATGAAAG	1260
	ATGGCGTCCT TAATCTTGAA AATGAAACAG TAGAATAGTT TTATTAAGCC GGTACATCAT	1320
25	GTGCCGGTAT TTTTATGTTT ATGTATTATT TGAATAAACT TTCACATTCA ATTAATAATA	1380
	ATTATTATCG AAAATCAGAA ATATTCCGTG AAATATAATA TTTTTGTAG TAAATGGCC	1440
	TCTAAGTATT CAATATTTAA ATATGGGGAT TGAATATAAA ATTATCGTAA TGGGGGTCAA	1500
30	TGGTTATGGA TTTATTGATA GGTACTTTAT TTTTATTTT GGTCTTAGTG ATTTTACAT	1560
	TATTTACATA TAAAGCGCCT AATGGTATGC GTGCCATGGG AGCATTAGCT AATGCAGCAA	1620
	TCGCAACATT TTTAGTGGA GCATTTAATA AATATGTTGG TGGCGAAGTA TTCGGTATTA	1680
35	AATTTTTAGA AGAGCTAGGA GACGCTGCGG GAGGTCTAGG TGGTGTGCT GCGCTGGAT	1740
	TAACAGCATT AGCTATCGGT GTGTCACCAG TATATGCATT AGTTATAGCA GCCGCTGCG	1800
40	GTGGTATGGA TTTATTACCA GGTTCCTTT CGGGTTATAT GATTGGATAT GTGATGAAAT	1860
	ATACAGAGAA ATATGTGCCG GATGGTGTG ACTTAATTGG ATCGATTGTC ATCTTAGCGC	1920
	CATTAGCTCG TCTTATTGCA GTATTATTAA CGCCAGTAGT GAATAGTACA TTGATTGAA	1980
45	TTGGTGATAT TATCCAAAGT AGTACGAATA CGAATCCAAT TATCATGGGT ATCATTTTAG	2040
	GTGGTATTAT TACGGTTGTC GGCACAGCGC CATTGAGTTC AATGGCATTG ACAGCATTAT	2100
	TAGGTTTAAC GGGTGTACCT ATGGCTATTG GTGCCATGGC AGCATTAGT TCGGCATTTA	2160
50	TGAATGGGAC GCTATTCCAT CGCTTAAAT TAGGTGATCG TAAGTCTACG ATTGCAGTAA	2220
	GTATTGAACC TTTATCACAA GCAGATATTG TATCAGCCAA TCCAATTCCA ATCTATATTA	2280
55		

## EP 0 786 519 A2

	ATGCGACAGG TACAGCTACA CCGATTGCAG GATTTT TAGT TATGTTTGA TTTAATCATC	2400
	CGACGACAAT TGTGATTTAT GGTGTAGTAA TGGCGATTGT AGGTGCGCTT GCAGGTTATC	2460
5	TTGGTTCAAT TGTATTTAAA AAATATCCAA TTGTTACTAA GCAAGACATG ATTAATCGAG	2520
	GTGCAGTAGA CGCATAGCAT CATCATATTG AATAGTAAAA ACAAATAAAA CATAGTAACG	2580
	TGATTCAGTC GATGTAACAG TCGATAATGA GTCACGTTTT TTTATAGAAA AATACAAGAC	2640
10	ATAAAAATGT CATAATTTAT TGTGACAAA TATCATACTG TATAAACATT TATCATTTTC	2700
	TCAAGTACCT TTTACACGAT GGAATGAACT TACTTTTAC GAAATTATGC GTATTTTATA	2760
	AACAAATATC ATTGATATAA CGGTAAATGT AAGCGTTTAC AACAGAAATA ACAGCATGCT	2820
15	ACGATATTTT TGTAAATTCA CTGATTCAAG TATTTTAAAGT CAATATGAGG AGGGATGTTA	2880
	TGAGCGATTG TGAGAAAGAA ATTTTAAAAA GAATTAAAGA TAATCCGTTT ATTTCAAC	2940
	GTGAACTTGC TGAGGCAATT GGATTATCTA GACCCAGCGT AGCAAACATT ATTCAGGAT	3000
20	TAATACAAA GGAATATGTT ATGGGAAAGG CATATGTTTT AAATGAAGAT TATCCTATTG	3060
	TTTGTATTGG CGCAGCGAAT GTAGATCGTA AGTTTTATGT GCATAAAAAAT TTAGTTGCAG	3120
25	AAACATCAAA TCCTGTAACG TCAACACGCT CTATTGGTGG CGTAGCAAGA AATATTGCTG	3180
	AGAACTTAGG TAGGCTTGGC GAAACGGTCG CTTTTTATC TGCTAGTGGA CAAGATAGTG	3240
	AATGGGAAAT GATTAAACGA TTGTCCACAC CATTTATGAA TTTGGATCAT GTTCAACAAT	3300
30	TTGAAAATGC GAGTACAGGT TCATATACAG CTTTAATTAG TAAAGAAGGC GACATGACAT	3360
	ATGGCTTgC AGATATGGAA GTGTTTGACT ACATTACGCC TGAATTTTTA ATTAAGCGTT	3420
	CACACTTATT GAAAAAGGCT AAGTGCATTA TTGTAGATTT GAATTTAGGC AAAGAGGCAT	3480
35	TAAACTTCTT ATGTGCCTAT ACCACGAAAC ATCAAATCAA ATTAGTTATC ACCACGGTTT	3540
	CTTCCCCAAA AATGAAAAAT ATGCCTGATT CATTACATGC TATTGATTGG ATTATCACGA	3600
	ATAAAGATGA AACAGAAACA TACTTAAATT TAAAAATAGA ATCTACTGAT GATTTAAAAA	3660
40	TAGCTGCTAA ACGCTGGAAT GATTTAGGTG TTAATAATGT TATTGTGACA AATGGCGTGA	3720
	AAGAACTCAT TTATCGAAGT GGTGAGGAAG AAATCATTA GTCAGTTATG CCATCAAATA	3780
45	GTGTGAAAGA TGTTACAGGT GCAGGCGATT CATTCTGTGC TGCAGTAGTG TATAGCTGGT	3840
	TAAATGGGAT GTCTACTGAA GATATATTAA TTGCTGGTAT GGTTAACGCA AAGAAAACGA	3900
	TAGAAACGAA ATATACAGTT AGGCAAAACC TAGATCAACA GCAACTTTAT CACGATATGG	3960
50	AGGATTATAA AAATGGCAAA TTTACAAAAG TATATTGAGT ATTCTCGAGA AGTTCAGCAA	4020
	GCACGGGAGA ACAATCAACC GATTGTAGCA TTAGAATCAA CAATTATTTT GCATGGTATG	4080

55

	GCCATTCCAG CAACCATAGC CATTATAGAT GGCAAAATTA AAATTGGTTT AGAAAGCGAA	4200
	GATTTAGAAA TACTGGCAAC TAGTAAAGAC GTTGCTAAAG TATCTAGAAG GGATTTAGCA	4260
5	GAAGTTATTG CGATGAAGTG TGTGGTGCT ACTACTGTAG CGACGACGAT GATATGTGCT	4320
	GCAATGGCTG GTATTCAATT TTTTGTACA GGAGGTATTG GGGGCGTCCA TAAAGGTGCA	4380
	GAACATACGA TGGACATTTC AGCAGACTTA GAAGAACTGT CTAAAACAAA TGTCACTGTT	4440
10	ATCTGTGCAG GTGCCAAATC AATTTTAGAC TTACCTAAGA CGATGGAGTA TTTAGAAACA	4500
	AAAGGCGTTC CAGTTATTGG ATATCAAACG AATGAATTGC CAGCATTCTT CACTCGCGAA	4560
15	AGCGGTGTTA AGTTAAACAAG TTCGGTTGAA ACGCCAGAAC GACTTGCTGA CATTCAATTA	4620
	ACAAAACAGC AGTTAAATCT TGAAGGTGGC ATTGTTGTG CTAAATCCAAT TCCATATGAG	4680
	CATGCCTTAT CAAAAGCATA TATTGAGGCA ATCATAAATG AAGCTGTTGT TGAAGCGGAA	4740
20	AATCAAGGTA TTAAAGGTAA GGACGCCACA CCGTTCTTGT TAGGGAAAAT TGTAGAAAAA	4800
	ACGAATGGTA AAAGTTTAGC AGCAAATATA AAAGTTGTTG AAAACAATGC GCGGTTGGGT	4860
	GCTAAAATTG CTGTCGCTGT TAATAAATTA TTGTAGGTGA TGATACATGA ATATTTTATT	4920
25	CGCTATCACA GGGATAGCAT TTGCACTATT TGTTCGTTT TTATTCAGTT TTGATCGTAA	4980
	AAAAATAGAC TTCAAAAAGA CGTTAATAAT GATATTTATT CAAGTGTTGA TCGTGTTATT	5040
	TATGATGAAC ACAACGATTG GTTTGACAAT TTTAACTGCA CTAGGTTTAT TTTTGAAGG	5100
30	GCTAATAAAT ATTAGTAAAG CAGGCATAAA TTTTGTTTT GGAGATATAC AAAATAAAAA	5160
	TGGCTTTACG TTCTTTTAA ACGTATTACT GCCATTAGTT TTTATTTCTG TATTAATAGG	5220
	CATCTTTAAT TATATTAAGG TATTACCATT TATTATCAAA TATGTAGGTA TCGCTATTAA	5280
35	TAAATAAAT AGAATGGGGC GCTTAGAAAG TTATTTTGCT ATTCAACAG CAATGTTTGG	5340
	GCAACCAAGG GTATATTTAA CAATAAAAGA TATTATTCCA AGATTATCTA GAGCGAAATT	5400
40	ATATACAATT GCGACGTCTG GTATGAGTGC TGTTAGTATG GCAATGCTAG GTTCATATAT	5460
	GCAGATGATT GAACCCAAGT TCGTAGTTAC AGCAGTAATG TTAAATATTT TTAGTGCGCT	5520
	TATCATCGCC AGTGTAATCA ATCCCTATAA ATCTGATGAT ACTGATGTTG AAATTGATAA	5580
45	CTTAACGAAA TCCACAGAAA CTAAACATT GAATGGAAAA ACAGGAAAAC CTAAGAAAGT	5640
	TGCCTTTTTT CAAATGATTG GTGATAGTGC GATGGATGGG TTTAAAATCG CTGTTGTAGT	5700
	AGCCGTAATG TTGTTAGCAT TTATTTTATT AATGGAAGCA ATTAATATCA TGTTTGGTAG	5760
50	TGTTGGTTTG AACTTTAAAC AGCTTATTGG CTATGTGTTT GCACCAATCG CATTCTTAAT	5820
	GGGGATTCCA TGGAGCGAAC TGTTCCAGCT GGCTCTTTAA TGGCGACTAA ATTAATTACA	5880
55		

	CAAGGTATCA TTTCAGTTTA CTTAGTAAGc TTCGCTAATT TTGGTACGGT TGGTATCATC	6000
	GTAGGTTCAA TTAAAGGCAT TAGTGATAAA CAAGGAGAAA AAGTTGCATC CTTTGCAATG	6060
5	AGGTTGCTAC TTGGTTCAAC TCTAGCTTCA ATCATTTCAG GATCAATCAT TGGCTTAGTA	6120
	TTGTAAATGA ATCGAAGTAC CTAAATTAAA TTCATGGCAA AGCTAAACCC CGTCACCAAG	6180
10	TTGGCGCAAC AGCGcATgcA TAACTTAGTG ACGGGGTTTT ATCATAACAA TCTACTTTTT	6240
	CGTAGCCGTT TTTGAAATGT ATGTTGATGG TTTATCTTTT TCAAAAATTG TTAATCCCGT	6300
	TATATCTTTT TTATGTTTTG AAGGGACAAT GAAGCTAAGT ATATAAGCAA AGACAAAAGC	6360
15	AACTGTAAAT GAAATGGTAG ATACATAGAA AGGTGAGTTA CCTTTGCCAA CACCATTATA	6420
	GACATAAGCA AAGATGATAC CCAATATTAA TCCACAAATA ACACCGAATG TATTCTGACG	6480
	TTTAGTGAAG ATACCAACTG CAAATACACC AGCCAATGGA ACGCCGAATA ATCCAGTCAC	6540
20	AAACAAGAAT AAATCCCATA AGTCATTGGA ATTAGAAGCA ATTAAGTATA GTGACATTCC	6600
	AAAACCGAAA ATACCTGCAA TGATAATAAT GAAACGTGCA AAGTTAACTT CGTGTCGCTC	6660
	GCTACCTTTT CCGAAGAAGC GTTGCTTAAT GTCGATTGAA ATACAAGCAG ATATAGAATT	6720
25	TAAACTAGAT GAAATGGTAG ACTGTGCAGC GCGGAAAATG GCTGCAATAA GTAATCCTGC	6780
	TACAAATGGT GGCATCTCAG TCAAAATGAA ATATGGCACT ACAGATGATG TATTGAAGCC	6840
	TTTTGGTAAA ACAGCTTCAT GTGTATAAAA TGAATACAGC ATTGTACCCA TACCATAAAA	6900
30	TAAGGGTGCT GAAATTAAAG CTAGGATACC ATTTGTCCAT AACGATTTAT TTGTTTCTTT	6960
	TAACTATCA GAAGCTTGAT AACGCTGCAC GACGTCTTGA CTCGCTGTGT ATTGATACAA	7020
35	GTTGTTGAAA ATATTTCCCTA GGAAAATAAT TGGAATGGCA GCTGCCGCAG TATTTAGTTT	7080
	CCAATTGTCT GCACTAATTA ATTTTTGTG CTCAATCGCA TCTGCAAAGA CAGTGCCGAA	7140
	ACCGCCTTTA ATGTTCAAA CACCTAGAAT AATAATAACT AAAGCGCCGC CTAATAAAAT	7200
40	GACGCCTTGA ATGAAATCAC TCCAAACCAC ACCTTCGAAA CCACCTAAAA ATGTATATAA	7260
	AATACATAGT AAACCAACGA GTGATGCAAC GATATAAGGG TTCATGTCTG ATACAGATGT	7320
	GATTGCTAAT GTTGTAAGT AGATAACAAAT TGCAACACGC CCTAAATGGT AAACGACAAA	7380
45	TAATAATGAG CCAATGACAC GTATGCTAGG GCCAAATCTA GCTTCTAAAT ATTCATATGC	7440
	AGATGTTACC TTAACTTTT TAAAGAAAGG GACATAGAAA TAAATAAGTA ATGGAATAAT	7500
	TGCGACGATA GCAATGTTAC CAGCGATATA TGACCAATCT GTTAAAAATG CTTTCTCTGG	7560
50	TGTCGACATA AATGTAATCG CACTTAACGT AGTAGCATAA ATTGAAAAGC CAACTACCCA	7620
	AGATGGCAAG CGACCACTTG CGGTAAAGAA ACTATTGGTA CTTTGGCTCG CGCGCTTGGT	7680

55

TGTGCCAAAT CCAACTTCTT TCATGGGCAA CATCCCCTTT ACAATGTATT GATTCTTTGA 7800  
 TGTCTATAAA TCGTATTTTG CAATGAGTTG ATCTAATGTT TGTCGATGTG CTTCGTTAAA 7860  
 5 AGGTTTGAAA GGTCTTTTCG GTAATCCTGC ATCAATGCCA CGATGACGTA ATATTCTTTT 7920  
 CAATGTTGGA TAAATCCCA TTGATAACAC TGTTCGATA ATGTCGTTTG AATCATGTTG 7980  
 CAGTTGGTAA GCTTCTTGAA TTTGACCTTG TCGTGCTAAG TCGAAGATTT TTCTTGACG 8040  
 10 GCGACCATTA ACGTTATATG TAGAACCAAT TGCACCATCT ACGCCAGAAA TCGTAGCTTG 8100  
 AACTAACATT TCATCAAAGC CAGATAAGAT TAATTTGTCT GGGAATGCTT TTCTAATACG 8160  
 TTCGAGTAGG AAGAAGTTTG GCGCTGTATA TTTAACACCA ACAATTTTTT CATGATTAAA 8220  
 15 TAGCTCGCTG AATTGTTCAA TAGAAATATT CACACCTGTT AAATCTGTA TTGCATAAAT 8280  
 AATCATATTG TTCTGAGTTG CTTGATAAT ATCGAAATAG TAATCTCTAA TTTCTTCAAA 8340  
 AGTAAATGGA TAGTAGAATG GTGTTACGGC AGAAAGTGCA TCATAACCGA GTTCTGTGGC 8400  
 20 ATATTTTCCA AGTTCAATGG CTTCAATTTAA ATCTAACGAA CCTACTTGAG CAATCAATTT 8460  
 CACTTTATCC CCAACTGCCT CTTTGCCAAC CTTGAAAAC TGTCTCTTCT GCTCTGTATT 8520  
 25 TAATAAAAAG TTTTCGCCTG AGCTACCATT TACATAAAGA CCGTCTAATT CTTCAAGTTTC 8580  
 AATGGCATT TTAGCAATTT GTTTAAGTCC TTGTTCAATT ACTTGACCAT TTTCATCAAA 8640  
 AGGAACGAGT AACGCTGCAT ATAAACCTTT TAAATCTTTG TTCATTATGA AGTCCCTCCA 8700  
 30 AAAATCATT GATAATATAG TTTACAGCTA TAATTGTAAA CGCTATCATA AAATGTAACA 8760  
 ATATCTTTTT GAAAATTGTA GTCATATTTA TGTATAATTA ATGAAAATGT TTTTCAAAT 8820  
 CAATAGAAAT GGAGTGAGTA AGGTGTATTA CATCGCAATC GATATTGGAG GCACTCAAAT 8880  
 35 TAAATCGGCA GTTATGATA AGCAATTGAA TATGTTTGAC TATCAACAAA TATCAACGCC 8940  
 GGACAACAAA AGTGAGCTTA TTAGTGACAA AGTATATGAG ATTGTAACAG GATATATGAA 9000  
 GCAATATCAG TTGATCCAAC CTGTCATAGG TATTTTCATCA GCAGGCGTTG TTGATGAACA 9060  
 40 AAAAGGCGAA ATTGTATACG CAGGGCCAAC CATTCCGAAT TATAAAGGTA CTAATTTTAA 9120  
 GCGATTATTA AAATCACTGT CTCCTTATGT CAAAGTAAAA AATGATGTAA ACGCTGCATT 9180  
 ACTAGGCGAA TTGAAATTAC ATCAATATCA AGCAGAACGG ATCTTTTGTA TGACGCTTGG 9240  
 45 TACAGGCATT GGGGGTGCCT ACAAGAATAA TCAAGGTCAT ATTGATAATG GTGAGCTTCA 9300  
 TAAGGCAAAT GAAGTTGGGT ATTTATTGTA TCGTCCAAC TAAATACAA CGTTTGAGCA 9360  
 50 ACGTGCTGCA ACGAGTGCAT TGAAAAAGCG CATGATTGCC GGAGGATTTA CGAGAAGCAC 9420  
 ACATGTGCCA GTATTGTTG AAGCAGCTGA AGAAGGTGAT GATATTGCAA AACAAATATT 9480

55

	AGGGCTTATA TTAATTGGGG GCGGTATATC TGAACAAGGA GATAATCTCA TTAAATATAT	9600
	CGAGCCGAAA GTTGCACT ACTTACCAA AGACTATGTT TATGCACCAA TACAAACGAC	9660
5	TAAGAGTAAA AATGATGCAG CATTATATGG CTGTTTGCAA TGATAGTTGA AAGAAGGAGT	9720
	CATTCTAAAA TAGAATTGTA AACCGTTACG AGAGATGAGA GCTGTTGTTA GTTCCACACA	9780
	TCACACTCTA TCTAGGACCA ATCTAAACTA TATCAACCAA CAGTGTGCCA CGGGCAAATT	9840
10	AAATGAAGA AGCTGAGATA TTAATAATTT AGAAAATGTA AAAAAATATT TGGTATTGAA	9900
	ATTAAAAAAG CACCTAGCAA CTCGTTGGGA CAATCACGAT GATTGTCTAC AGTTGCAGGT	9960
	GGATTTGAAT ATACTACTAG TTATTTGTTG TCTAGGATAA TAGATTAGT ATGTTGATAA	10020
15	GTTTGACTCA GATTCGTATT TTCTAATAAA TGATAACTCA CGATATCGAT TAAAAAGAGT	10080
	GTCGCAATTT GTGTGTTGAT AAATTGATGG TCGGTATTAC GCGATTGATC CGTTGTTAAA	10140
20	AGTACTAAAT CTGCACAATC TGTAAGTTTA CTACCTTCAA AATTTGTGAT GGCAACGACA	10200
	TATGCACCAT GAGATTTGGC GACTTCCGCT GCAGAAATTA ATTCCGAAGT ATTACCACTA	10260
	TTTGACATAG CAATAAACAT ATCCGAATGA GATAGTAGGG ATGCCGATAT TTTCATTAAA	10320
25	TGTGAATCGG TAGTAACATT ACCTTTTAGC CCCATACGAA TCATACGATA ATAAAATTCA	10380
	GTCGCTGATA AACCAGAGCT ACCTAGTCCA GCAAAGAGTA TATGTCGACT TGATTGAAGT	10440
	TTGTCGATAA AGGTTTGAT AATGTCGTTA TCAATAAATT CACCAGTTTG TTGAATGATT	10500
30	TGTTGATGAT ATTTATGAAT TCTTTGAATA ATTGGGCTAT TTTCAATAAC TGTCTCTGTC	10560
	ATTTCTTGTT GAATATTAAA TTTTAAATCT TGGAAATCT CATAATCCAG CTTATGACTA	10620
	AAGCGTGTCA TCGTTGCTGG TGATGTACCA ATCGCATGGG CTAAGGAGTT AATCGTTGAA	10680
35	AAGGCATCGC TATAACCATT TTGTCTTATA TAATTGACGA TCGTTTATC AGTTTTTGTA	10740
	AATAAATGTT GATAACGTTG AACACGATTC TCAAATTTCA TTGTGTCACC CCTTCATCTT	10800
40	AATGATTACT ATTATATATG AAAAAATTT TCAAGATAGT AAAAAGCATT GATAAAAATT	10860
	ATCTTAATGA TATATTGTAA ATGACTTTAC GTGAAAAAC GACTTATGGA GTGAGGAATA	10920
	ATGTTACCAC ATGGATTAAAT AGTATCTTGT CAGGCACTAC CAGATGAACC ATTGCATTCA	10980
45	TCTTTTATTA TGTCGAAAAT GGCATTAGCT GCGTATGAAG GTGGTGCTGT TGGTATTCGC	11040
	GCAAATACTA AGGAAGACAT TTTAGCAATT AAAGAAACGG TAGATTTACC AGTTATTGGC	11100
	ATTGTGAAAC GTGACTATGA TCACTCAGAT GTTTTCATTA CTGCAACGTC AAAAGAAGTT	11160
50	GATGAACTGA TAGAAAGCCA ATGTGAAGTC ATTGCATTGG ATGCAACGTT ACAGCAACGT	11220
	CCGAAAGAAA CGTTAGACGA ATTAGTATCA TATATTAGAA CACATGCACC GAACGTTGAA	11280

55



EP 0 786 519 A2

	TATATTGGCA CGACGTTACA TGGCTATACT AGTTATACGC AAGGACAATT ACTTTATCAA	11400
	AATGACTTCC AATTTTTTAAA AGATGTACTA CAAAGTGTTG ATGCAAAAGT TATTGCGGAA	11460
5	GGTAATGTCA TTACACCGGA TATGTATAAA CGTGTGATGG ACTTAGGCGT TCATTGTTCA	11520
	GTCGTTGGTG GTGCGATAAC ACGACCAAAA GAAATTACGA AACGTTTTGT TCAAATTATG	11580
	GAAGATTAAA TGATAACGAT AAAAAACGA GATGACCATC ATTAATTAAA GGCACCTAAT	11640
10	TATCTTAGGT GGCTGAATGA ATGTAATGGG TTCATCTCGT TTTGTTTGTT TATGATAGTG	11700
	ATTTTATTTT CAACTTTATC CAAAAATAAG TAAAGCGACG GGGATGGTGA TTAATAGCGA	11760
15	CAACGCCACG CGTAAAAACC AAATGATGAT GAGTTTCCAG ACAGGTATTT TAATTTTCAGT	11820
	TGCTAGTATA CATGGCACTA ATGCTGAGAA AAAGATAATG GCTGATACGC TTAATACACC	11880
	GACGACAAAT TTAGTACTCA TTGCAGCTTT AGTTACTAAC AAAGATGGTA GAAACATCTC	11940
20	TACAATAGAA AaCTGACGC TTTTGCTAGT AAAGCCTGAT CAGCAATTGG GAAAATATAA	12000
	ATAAATGGAT AGAAGATATA GCCAAGCCAA TCAATGAATG GTGTATAGTT CGCTACAATC	12060
	AGTCCTAAAA AACCAATCGA TAATATAGAA GGTAAAATAC CAACAGTCAT TTCTAAACCG	12120
25	TCTTTCAAAT TGTCCCAAAC GTTCTTCACG AGAGATGGTG TTAATGCATT TTGTTTCATC	12180
	GCCTCTGCAT ATGCAGTTTT CAGTCTGCTT CCTTCAATAG CAACTTCTTG TTCTCCTTCT	12240
	TGTCGGTTAT AATATTCTGT TGATTCAATG CTGATTGGCG GTAGCCATGC AGTAATTGCA	12300
30	GTCACGACAA ATGTGATGAC TAAAGTTATC CAAAAGTATA AATTCCAATG CGGCATTAAT	12360
	CCTAAAGTTT TAGCAACGAT AATCATAAAA GTTGCTGAAA CTGTTGAAAA GCCAGTCGCA	12420
	ATAATCGTGG CTCTCGTTT GTTGACATC CCTTGCTTAT AGACACGATT AGTAATCAAT	12480
35	AATCCTAAGG AATAACTGCC GACAAACGAA GCCACTGCAT CGACAGCGGA TTTTCCTGGT	12540
	GTTTTAAAAA TAGGTCTCAT AATAGGCTCC ATATAAACAC CGACAAATTC TAATAAGCCA	12600
40	TAGCCCACTA ATAAAGAAAG cGcAATTGCA CCTACTGGAA TTAAGATACT TAATGGCATC	12660
	ATTAATTTTT CAAACAAAAA CGGACCATAG TTAGCTTTAA ATAGTATTGA TGGACCGATT	12720
	TTAAATACAT ACATTATACC GATCATTGCA CCTGCAACTT TAAATAATGT AATGACCAAG	12780
45	TTTGTGATTG AAGTCATAAA AGTACGTCTC ACTATTGGTA ACGCTGTACC AATTAAAATC	12840
	ATAATCAGTG CAACATAGGG CATAAGTGA CCTATGATTG AGCGAATGGC TAGATGAACA	12900
	TGATCGACGA AAATAGTGTT GTTACCATTA ATCGTAAAAG GAATAAAGAA ACATAGTATG	12960
50	CCCACTAAAC TATAGACAAA AAAACGCCAT GCACTTGGTT GTTGTGCATT AGAATGATAT	13020
	TGATTCATTA AAGCAACCCC TTTGTTTAAA TGAATACACA AACTGTATG ATGCATCTTC	13080

55

ATAGTTTGAA TTATTTTCAT ACCAATACAA ATTAACATA TATATATAGA TTGAACTAT 13200  
 ATTACTTAAT AAAATATTTA TCTTAAATGT TGTGTGTTG ATTCAACACC ACAACTAAAA 13260  
 5 GTGTTTATAA ATTATTTGGA AATACACATA TTTGTAAATG ATTAGTATCG ATTTAATATC 13320  
 GTATTATTAA ATTTTATTAT ATTTTGTAGT CTTAATCMAA AAATAATATA TGTCATGTTA 13380  
 TATTGAAGGT GCAGTTGTTT TTCATTCTCA AGAGGGGGTC AAAAAAATAC TTTTGAGGTG 13440  
 10 ATTATATGTT AAGAGGACAA GAAGAAAGAA AGTATAGTAT TAGAAAGTAT TCAATAGGCG 13500  
 TGGTGTCACT GTTAGCGGCT ACAATGTTTG TTGTGTCATC ACATGAAGCA CAAGCCTCGG 13560  
 AAAAAACATC AACTAATGCA GCGGCACAAA AAGAAACACT AAATCAACCG GGAGAACAAG 13620  
 15 GGAATGCGAT AACGTCACAT CAAATGCAGT CAGGAAAGCA ATTAGACGAT ATGCATAAAG 13680  
 AGAATGGTAA AAGTGAACA GTGACAGAAG GTAAAGATAC GCTTCAATCA TCGAAGCATC 13740  
 AATCAACACA AAATAGTAAA ACAATCAGAA CGCAAAATGA TAATCAAGTA AAGCAAGATT 13800  
 20 CTGAACGACA AGGTTCTAAA CAGTCACACC AAAATAATGC GACTAATAAT ACTGAACGTC 13860  
 AAAATGATCA GGTTCAAAAT ACCCATCATG CTGAACGTAA TGGATCACA TCGACAACGT 13920  
 25 CACAATCGAA TGATGTTGAT AAATCACAAC CATCCATTCC GGCACAAAAG GTAATACCCA 13980  
 ATCATGATAA AGCAGCACCA ACTTCAACTA CACCCCGTC TAATGATAAA ACTGCACCTA 14040  
 AATCAACAAA AGCACAAGAT GCAACCACGG ACAACATCC AAATCAACAA GATACACATC 14100  
 30 AACCTGCGCA TCAAATCATA GATGCAAAGC AAGATGATAC TGTTCGCCAA AGTGAACAGA 14160  
 AACCACAAGT TGGCGATTTA AGTAAACATA TCGATGGTCA AAATTCCCCA GAGAAACCGA 14220  
 CAGATAAAAA TACTGATAAT AAACAACATA TCAAAGATGC GCTTCAAGCG CCTAAAACAC 14280  
 35 GTTCGACTAC AAATGCAGCA GCAGATGCTA AAAAGGTTTG ACCACTTAAA GCGAATCAAG 14340  
 TACAACCACT TAACAAATAT CCAGTTGTTT TTGTACATGG ATTTTLAGGA TTAGTAGGCG 14400  
 ATAATGCACC TGCTTTATAT CCAAATTATT GGGGTGGAAG TAAATTTAAA GTTATCGAAG 14460  
 40 AATTGAGAAA GCAAGGCTAT AATGTACATC AAGCAAGTGT AAGTGCATTG GGTAGTAAGT 14520  
 ATGATCGCGC TGTAGAACTT TATTATTACA TTAAAGGTGG TCGCGTAGAT TATGGCGCAG 14580  
 CACATGCAGC TAAATACGGA CATGAGCGCT ATGGTAAGAC TTATAAAGGA ATCATGCCTA 14640  
 45 ATTGGGAACC TGGTAAAAAG GTACATCTTG TAGGGCATAG TATGGGTGGT CAAACAATTC 14700  
 GTTTAATGGA AGAGTTTTTA AGAAATGGTA ACAAGAAGA AATTGCCTAT CATAAAGCGC 14760  
 50 ATGGTGGAGA AATATCACCA TTATTCATG GTGGTCATAA CAATATGGTT GCATCAATCA 14820  
 CAACATTAGC AACACCACAT AATGGTTCAC AAGCAGCTGA TAAGTTTGGA AATACAGAAG 14880

55

EP 0 786 519 A2

	ATTTAGGATT AACGCAATGG GGCTTTAAAC AATTACCAA TGAGAGTTAC ATTGACTATA	15000
	TAAAACGCGT TAGTAAAAGC AAAATTTGGA CATCAGACGA CAATGCTGCC TATGATTTAA	15060
5	CGTTAGATGG CTCTGCAAAA TTGAACAACA TGACAAGTAT GAATCCTAAT ATTACGTATA	15120
	CGACTTATAC AGGTGTATCA TCTCATACTG GTCCATTAGG TTATGAAAAT CCTGATTTAG	15180
	GTACATTTTT CTTAATGGCT ACAACGAGTA GAATTATTGG TCATGATGCA AGAGAAGAAT	15240
10	GGCGTAAAAA TGATGGTGTC GTACCAGTGA TTTCGTCATT ACATCCGTCC AATCAACCAT	15300
	TTGTTAATGT TACGAATGAT GAACCTGCCA CACGCAGAGG TATCTGGCAA GTTAAACCAA	15360
	TCATACAAGG ATGGGATCAT GTCGATTTTA TCGGTGTGGA CTTCTGGAT TTCAAACGTA	15420
15	AAGGTGCAGA ACTTGCCAAC TTCTATACAG GTATTATAAA TGAATTGTTG CGTGTGAAG	15480
	CGACTGAAAG TAAAGGAACA CAATTGAAAG CAAGTTAAAT TCATCTTCTG AATTTAATAT	15540
	GCTATGTAAA TCGTGCTGTT ATCATGGCAC ATCAGATATA AGTAGCATCA CAGTGTGAA	15600
20	TTTAAAAATA GTAAAGTGAA ATAAAGCGCC TGTCTCATTA GCGAAACTA AAGGGACAGG	15660
	CGTATCTGTT TATGAGCTTA ATAAATTGTA TGAATAATAT GGTGATCGA ATAAGTGT	15720
25	ATCATGATGA TAAATTGAGT TTTTAAAAAT AATGATATAT TACATCAITG TTATAGCGTT	15780
	TAAGAAATCA ACAACTTTAC GATAAATAGT GATTGCTTCG TCATTAGGTC TACGATCAA	15840
	ATCATGCTCG TTTTATTCA CGCGTTCAAA TGTGAATGT GGAACATGAT TCATGATATG	15900
30	TTGCTTTTCC TCAACGGGAA CATCATAATC GCCATTACAA TGCGCAATGA AAACAGGTGG	15960
	AAGTGTTTTA AGTTCATCTG GTGCAATATT ATATTTTGAA TTAGTATAAT CAGCAATGTT	16020
	AATCATATTT ATCCATTTAC CTGTGCCACG TGCAATAACG TAGATTAAAA AACGTTGTGC	16080
35	GATTTGATCT TGAACAACCG GTGTTGGTGA AGTGAGTTGT GCAATCAITG TTTCGTTTAC	16140
	GCTTTGAGCT ATTTTGGCGT AATAACTATT AGTTGTTTTA AAAGGTTTCA TGTGATGCG	16200
40	ACTATAACCA TAAAAATCAA TAACACCATC AATATCTCTG TCTCGTGCAA TTAATAGACT	16260
	TAAATATGCA CCTGATGATC TGCCAAAGGT AAAAATAGGG CAATTAGAAT ATTGTGATTG	16320
	AATCGCATCG AATGATGCgn AGnACATCCT CAATAATGCA ATCGAGACTT ACTTCTGGTA	16380
45	ATAAACGATA ACTTAGTTGA ATTAAATCGT AATGTTCCGT AAgATATCGA TATACTGTGG	16440
	GGATAAATCG TTAGCTTTAC CGAACATTAA TCCACCACCG TGGATGTAGA CAATAGCGCC	16500
	TTTGTGTTG TGATTTTTTG CTTTAATAAT TGTGTAAGGT AATGCAAATG CATCTTTAGT	16560
50	AATTACTTTA TCTTTAATTT CAGTCACGAT TTAATAGGCT CCTTATTTTT GATATTGATG	16620
	TCATTATAAC ACTGTCTTAA ATTTCCATGA AAAATAGTCT TAAGACGATG AGTCATGATA	16680

55

CATCATTTTA ACAATATCTT TAAAAGCAGC ATGTGGAATG GCTAAATCTT CTAAATCTGC 16800  
 CATAGAAAAT TCAAGATTGA TATCATGTGG TCGCTGTTCA GCAAGTTTAT GCACAAAGTC 16860  
 5 AGGTTCTGTG ACAAAGGCG AAGACATGCC GACCATATCT GCATGTTGTA AAGCATCTAA 16920  
 AGCAGACTCT GGAGAATTAA TCCCGCCACT TGCAATTAAA GGGATACGAC CTGCTAAATG 16980  
 10 TTCATAGACA ATTTGGTTAA CTGGTCGACC GAAATGATCA CCTGGTGATC GAGACGTATT 17040  
 TTGATAAATA TGTCGACCCC AGCTAGCGAT TGCTAAGTAT TGGATGTTTG AAACGTCCAT 17100  
 GACCCAATTG ATTAATTGGT TGAACGTCGTC AATGGTATAT CCTAAATCAC TGCCTCTGGT 17160  
 15 TTCTTCTGGC GTTGCTCGAA ATCCTAAAAT AAAATGTCA GGTGCTTCTT TATCAATCAC 17220  
 TTCTTGTAAC GCACGCATAA CTTCTAAACA TAATCTTGCA CGATTTTTTA ATGAGTCGGC 17280  
 ACCGTAATGG TCTGTACGTT TATTCGAAAA AGTTGAGAAA AATGTTTGAA TCAGCAAACG 17340  
 20 TTGTGCAATC GAAATTTCCA CACCATCAAA ACCTGCTTTA ATCGCGCGTA ATGTAGCATC 17400  
 GCGATACTGC TGAATGATGC TATTGATTTT CTCATGAGAC ATGGCGATAA CATCGTGTTT 17460  
 AATCGGTGAA TGCAATGTCA TAGGGCTTGG TCCATACACC TTTCCAAAAT TTAAAATGGC 17520  
 25 TTGAATTGAA AAACGACCAG CATGCGCTAG CTGGATAATA GCGAGGCTAC CATGTTGTTT 17580  
 CATCGTAGAT GCCATGTTAG TTAATCCAGG GATACAAGCA TCATGATCAA TATTAAAGCC 17640  
 ATATTCAAAC AATTGACCAT AAGGTTCAAT GTAAGCAGCG CCGGTGACTT GCATTCCAGC 17700  
 30 TGAATTAGAG CGACGTGCAG CATAAGCCAA GTCTTCTTTT GTAATATAGC CTCTTTTGT 17760  
 TGATGTGTTT ACGGTCATTG GTGATAATAC AAAGCGATTG GAAATTTTGA TGCCATTAGG 17820  
 TAAGTGGATT GATTGTAAAA GTGGTTTGTA TCGGTACATA CTATGATTCC TTTTCTATT 17880  
 35 AATATTGTTT TCAAAGTACC ATGGAAAGAA TGAATAATCA ATGATGAACA GTCTTGATAG 17940  
 AATAGAATTG GTACATGGAA AGTATTTTTA AAATTAACT AATGAATGGC ATTTGTAGGT 18000  
 CTGAAAATAT GAATATGAAA AAGAAAAATA AAGGCGAAAA GATATAAAAG TTAATTGAAA 18060  
 40 AACGTTATCA TATACGTGGG TATATGAAGA GGAATGGTA TTAAGAACGC TAAAATGTTA 18120  
 TGTCGGTTTG ACATGACAGG ATAAGTTTGG AGATGACGGA TTGGTTAAAT TAAGCGTATT 18180  
 AGACTATGCC TTAATAGATG AAGGTAAGGA TGCACAAAAG GCATTGCAAG ATTCAGTGAC 18240  
 45 ACTTGCAAAA TTAGCAGATC GACTTGGCTT TAAGCGAATT TGGTTTACGG AACATCATAA 18300  
 TGTACCAGCG TTGCGTGTA GTAGTCCAGA ACTTTTGATG ATGCATACAT TGGCGCAGAC 18360  
 50 AAATCACATA CGAGTTGGCT CTGGTGGTGT GATGCTGCCG CACTATCGAC CTTATAAAAT 18420  
 TGCTGAGCAT TTTAGAATGA TGGCAGCGTT ATATCCAAAT CGTATTGATT TAGGTATTGG 18480

55

EP 0 786 519 A2

	TAGTTACGAT GAATCGATTT CGTTATTACG TGATTATCTT ACAATAAAGG ATAAACCAAG	18600
	TGCGCATACG TTAGGTGTCC AACCACACAT TGATCATTTT CCAGAAATGT GGTTATTAAG	18660
5	TAGTAGCGCA ACATCTGCCA AAATAGCTGC CGAACTAGGT ATAGGGCTTT CTGTTGGAAC	18720
	ATTTTGTCTA CCAGATATAA ATGCGATACA TACAGCGAAG GATAACATTG ATATTTACAA	18780
	AAAACATTTT CAAGCATCAA CGATTAAAAT GGACGCAAAG GTGATGGCAT CTGTATTTGT	18840
10	CATTGTAGCT GATAACGAAG CGGAAGTAGC AGCATTACAA CATGCCTTAG ATGTTTGGTT	18900
	ATTAGTAAA TTACAATTTG CAGAATTTGA AGATTTTCCT TCAGTAGACA CAGCACAAAA	18960
15	GTATAAGCTT AATGATCGAG ACAAAGAGAT GATTCAAGCA CATCAAGCAC GCATCATTGC	19020
	AGGTACACAA GAAAAGGTTA AAGCACAATT AGATGATTTT ATTGCTACGT TTGAAGTTGA	19080
	TGAGGTGTTA GTAGCACCGC TTATTCCAGG TATTGAACAG CGTTGTAAAA CATTAAAATT	19140
20	ACTCGCGGAA ATTTATTTGT AGCATTTTAA ATAGAAGAGA AAGGATGAAG ATAAGATGAA	19200
	AAAGTTAGCC AATTATTTAT GGGTAGAAAA AGTAGGAGAT TTGTATGTGT TTAGTATGAC	19260
	ACCTGAATTG CAAGATGATA TTGGGACAGT AGGTTATGTT GAATTCGTAA GTCCAGATGA	19320
25	AGTTAAAGTG GATGATGAAA TTGTGAGTAT CGAAGCATCG AAAACGGTCA TTGATGTGCA	19380
	AACGCCATTG TCAGGAACGA TTATTGAGCG AAATACAAAA GCGGAAGAAG AACCGACAAT	19440
	TTTAACTCT GAAAAACCAG AAGAAAATTG GTTGTTCAAA TTGGATGATG TCGATAAAGA	19500
30	AGCATTCCTA GCATTACCGG AGGCTTAAAT GGAAACGTTA AAATCAAATA AAGCGAGACT	19560
	TGAATATTTA ATCAATGATA TGCATCGAGA GAGAAATGAC AATGACGTAT TGGTAATGCC	19620
	ATCTTCATTT GAAGATTTGT GGGAATTATA TCGAGGCTTA GCAAATGTCA GACCGGCATT	19680
35	ACCTGTAAGT GATGAATATT TAGCTGTACA AGATGCTATG TTAAGTGATT TGAATCGTCA	19740
	ACATGTTACG GATTTGAAGG ATTTGAAGCC GATAAAAGGT GACAATATCT TTGTTTGGCA	19800
40	AGGTGATATC ACGACGTTAA AAATCGATGC TATTGTAAAT GCTGCAAATA GTCGTTTTCT	19860
	AGGATGTATG CAAGCTAATC ATGACTGCAT TGATAATATT ATTCATACAA AAGCGGGTGT	19920
	TCAAGTTCGA CTTGATTGTG CAGAGATCAT TCGACAACAA GGGCGCAATG AAGGTGTAGG	19980
45	TAAAGCCAAA ATAACACGTG GATATAATTT GCCAGCAAAG TATATAATTC ATACGGTTGG	20040
	TCCGCAAATA CGTCGATTGC CTGTTTCAAA GATGAATCAG GACTTGTTAG CTAAATGTTA	20100
	TCTTAGCTGT CTTAAATTGG CTGATCAACA TAGTTTAAAT CATGTCGCTT TTTGCTGTAT	20160
50	ATCTACAGGT GTATTTGCTT TTCCTCAAGA TGAAGCAGCA GAAATTGCTG TTCGAACAGT	20220
	AGAAAGCTAT CTCAAAGAAA CAAATTCAAC ATTGAAAGTC GTGTTCAATG TATTTACAGA	20280

55

	CAATGTCTCT GTTAATGGAT GACAAGACAA AGCAGGCTGA AGTATTGCGT ACTGCGATTG	20400
	ATGAAGCAGA TCGGATAGTG ATTGGAATTG GTGCAGGCAT GTCTGCATCT GACGGATTTA	20460
5	CATATGTAGG AGAGCGTTTT ACGGAAAATT TCCCAGATT TATTGAAAAA TATCGCTTCT	20520
	TTGATATGTT GCAAGCGAGT TTACATCCTT ATGGCAGTTG GCAAGAGTAT TGGGCATTTG	20580
	AGAGTCGTTT TATTACATTA AACTATTTAG ATCAACCTGT AGGTCAGTCT TACCTCGCTT	20640
10	TAAAATCCTT GGTGGAAGGT AAACAGTACC ACATTATAAC TACGAATGCA GATAATGCTT	20700
	TCGATGTAGC TGATTATGAT ATGACTCATG TATTTTCATAT ACAAGGGGAG TATATACTGC	20760
	AACAGTGTAG CTCAGCATTG TCATGCTCAA ACGTATCGCA ATGATGATTT AATTCGTAAA	20820
15	ATGGTTGTTG CGCAACAAGA TATGCTTATA CCTTGGGAGA TGATTCCAAG ATGTCCAAAA	20880
	TGTGATGCCC CAATGGAAGT GAATAAACGT AAAGCGGAAG TTGGGATGGT TGAAGATGCT	20940
20	GAATTTTCATG CGCAACTACA TCGTTATAAT GCTTTTCTAG AGCAACATCA AGATGATAAA	21000
	GTGTTGTATT TGGAAATTGG AATTGGTTAT ACTACACCAC AATTTGTGAA GCATCCTTTT	21060
	CAGCGTATGA CACGTAAAAA TGAAAATGCC CTTTATATGA CGATGAATAA AAAGGCATAT	21120
25	CGCATTCGGA ATTCAATTCA AGAACGTACC ATACATTTAA CTGAGGATAT CTCAACATTG	21180
	ATTACAGCAG CACTCCGGAA CGACAGCACA ACGAAAAATA ACAACATTGG AGAGACAGAA	21240
	GATGTACTTA ATAGAACCGA TTAGAAATGG AGAATATATT ACTGATGGTG CGATTGCACT	21300
30	CGCTATGCAA GTTTATGTTA ACCAGCATAT CTTTTTAGAT GAAGATATTT TATTCCTTA	21360
	TTATTGTGAT CCAAAGTGG AAATTGGACG TTTTCAAAT ACTGCTATAG AAGTGAATCA	21420
	AGATTATATA GATAACACA GTATTCAAGT AGTTCCCGA GATACTGGTG GTGGCGCTGT	21480
35	GTATGTTGAT AAAGGTGCCG TTAATATGTG TTGTATTTTA GAACAAGACA CTTCAATTTA	21540
	TGGTGAATTT CAACGATTTT ATCAACCAGC TATAAAGGCG TTGCATACAT TAGGTGCAAC	21600
40	AGATGTGGTA CAAAGCGGTA GAAATGATTT AACATTGAAT GGTAAAAAAG TGTCAGGCGC	21660
	CGCAATGACA TTAATGAATA ATCGTATTTA TGGCGGTTAT TCGCTATTAC TTGATGTTAA	21720
	TTATGAAGCA ATGGATAAAG TGTAAAGCC TAATCGCAA AAGATTGCAT CGAAAGGGAT	21780
45	TAAATCTGTG CGCGCACGTG TTGGTCATCT TAGAGAAGCA CTGGATGAAA AGTATCGTGA	21840
	TATAACCATT GAAGAATTTA AAAATTTAAT GGTGACGCAG ATTTTGGGAA TCGATGACAT	21900
	TAAAGAGGCG AAACGATATG AATTAACGGA TGCAGATTGG GAAGCGATTG ATGAATTAGC	21960
50	TGATAAAAAG TATAAAAATT GGGATTGGAA TTATGGCAAG TCACCCAAAT ATGAATACAA	22020
	TCGAAGTGAA AGATTATCTT CAGGTACGGT AGACATAACA ATTTCTGTTG AACAAAATCG	22080

55

AGAAGCATT A CAAGGAACAA AAATGACAAG AGAAGATTTA ACGCATCAGT TAAAGCAATT 22200  
 AGACATCGTT TATTATTTTG GCAATGTTAC GGTAGAAGCA TTAGTGGATA TGATTTTAAG 22260  
 5 TTAATATTGT TATTTTATGT ATGCTGAATC ATTGGAAGTG TTTGCTTGCT CTTGAAAAGG 22320  
 TGACAATAGT GTTTGGTGAA GGTGAACAT ATGAGTGGAA ATTATTGCCT TTAAGTATTC 22380  
 AAAGTATGAT ATATATATGG TTTTGTTC TAAATGATTG GGTATTTGAA AATAGATGAG 22440  
 10 TTTAATATTT TAAGGAATAT AATGATGTTT ACTTTTATAA TTCATATAGA ATATTAAGCA 22500  
 ATATAAGTCT GTTGATATAT ACAAATATA ATGACTGCTA TAATGAGTAA TCAATAGACA 22560  
 CAAAGAGGAG ATTATGTGAT GAATAATAAA GTATTAGTAA CCGGTGGTAC AGGGTTTGTT 22620  
 15 GGCATGCGAA TTATTTACG ATTATTAGAA CAAGGTTATG ACGTACAAAC GACGATACGT 22680  
 GATTTAAGTA AAGCTGATAA AGTAATTAA ACAATGCAAG ACAATGGCAT TTCCACAGAG 22740  
 CGATTAATGT TTGTCGAAGC GGATTTATCA CAAGATGAAC ATTGGGATGA AGCAATGAAA 22800  
 20 GATTGCAAGT ATGTCTTGAG TGTAGCATCT CCGGTGTTTT TCGGTAAAAC AGACGATGCA 22860  
 GAAGTGATGG CGAaCTGcAA TTGAAGGTAT ACAACGTATT TTAAGAGCTG CAGAACATGC 22920  
 25 GGGTGTAAA CGTGTGGTAA TGAAGGTAT ACAACGTATT TTAAGAGCTG CAGAACATGC 22980  
 TAAAAATTCA ATCACAAATG AAAGTCATTG GACAAATGAA GATGAACCAG GCTTATCAGT 23040  
 ATATGAAAAA TCAAAATTGT TAGCTGAAAA GGCAGCGTGG GATTTTGTG AGAATGAAAA 23100  
 30 TACAACAGTA GAATTTGCCA CAATCAATCC AGTTGCAATT TTTGGCCAT CATTAGATGC 23160  
 ACACGTTTCA GGAAGCTTTC ATTTATTAGA AAATTTATTG AATGGTTCAA TGAAACGTGT 23220  
 ACCGCAAATT CCGTTAAATG TTGTTGATGT GAGAGACGTA GCTGAACTGC ACATTTTGGC 23280  
 35 AATGACAAAT GAACAAGCTA ATGGCAAGCG ATTTATTGCG ACGGCTGATG GACmaATTwa 23340  
 tTTGTTGGGA ATTGcCAAAt TAATTAAAGA AAAGGGCCTG GAAATAGCTC CAAAAGTTCC 23400  
 40 TACTAAAAAA TTACCCAGCT TTATTTTGAG CnAnGnGCC 23439

(2) INFORMATION FOR SEQ ID NO: 39:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4522 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCCTTTGAGA GTATATCATC TAGTCAAATT ATGCCTGTCA TTAGAGCGAC TAGCTTTGAT 60

	TATTATGCAG TCGATTTAGG GAAATCATAT CGTCTAATTG ACGAAAGCAT GTTAGAGGAT	180
	TTGAAGTTAA CTGAACAACA AATAAGAGAA ATGTCTCTGT TTAATGTTAG AAAATTGTCA	240
5	AATTCATATA CGACTGATGA AGTAAAAGGT AATATTTTTT ATTTTATTAA CTCAAATGAC	300
	GGGTATGATG CAAGTAGGAT ACTAAATACT GCATTTTTTAA ATGAAATTGA GGCACAATGT	360
	CAAGGCGAAA TGCTCGTAGC AGTGCCACAC CAAGATGTGT TAATTATTGC AGATATACGC	420
10	AATAAACAG GATATGATGT GATGGCACAT TTAACAATGG AATTTTTCAC TAAAGGTCTA	480
	GTTCGAATTA CATCATTATC CTTTGGATAT AAACAGGGTC ATCTTGAACC GATATTTATT	540
	TTAGGTAAAA ATAATAAACA AAAAGAGAT CCAAACGTGA TTCAGCGTTT AGAAGCAAAT	600
15	CGTCGTAAAT TTAATAAAGA TAAATAGAAA TAATTGGATA AGGAGTTTTC TCATAATGAA	660
	TTTATTTTAC AATCCTAAAT ATGTAGGAGA TGTCGCATTT TTACAAATTG AACCAGTTGA	720
	AGGTGAATTA AACTACAATA AAAAAGGTAA TGTGTGTGAA ATTACTAATG AAGGTAATGT	780
20	TGTAGGTTAT AATATTTTTC AAAATTCAAA AGATATAACA ATTGAAGAAA AAGGTCATAT	840
	TAAATTAAC TATGAACCTG TAAATGTATT CCAAAGCGT ATTTCAGAAG CTGGTTTTGA	900
25	TTATAAATTA AATGCTGATC TATCACCGAA ATTTGTAGTT GGCTACGTTG AACTAAAGA	960
	CAAACATCCT GATGCAGATA AATTAAGTGT ACTAAATGTA AACGTTGGAA ATGACACATT	1020
	ACAAATTGTA TGTGGCGCGC CTAACGTTGA AGCTGGACAG AAAGTTGTTG TTGCTAAAGT	1080
30	AGGTGCAGTG ATGCCTAGCG GTATGGTAAT TAAAGATGCT GAATTACGTG GTGTTGCCTC	1140
	AAGCGGTATG ATTTGTTCAA TGAAAGAATT GAATTTACCT AATGCACCTG AAGAAAAAGG	1200
	TATTATGGTA TTAATGACA GCTATGAAAT TGGACAAGCA TTCTTTGAAT AATTAAGGAA	1260
35	GGTAGTGAAA ATATGAGCTG GTTTGATAAA TTATTCGGCG AAGATAATGA TTCAAATGAT	1320
	GACTTGATTC ATAGAAAGAA AAAAGACGT CAAGAATCAC AAAATATAGA TACGATCAT	1380
	GACTCATTAC TGCCTCAAAA TAATGATATT TATAGTCGTC CGAGGGGAAA ATTCCGTTTT	1440
40	CCTATGAGCG TAGCTTATGA AAATGAAAT GTTGAACAAT CTGCAGATAC TATTTAGAT	1500
	GAAAAAGAAC AATACCATCG AGACTATCGC AAACAAAGCC ACGATTCTCG TTCACAAAAA	1560
	CGACATCGCC GTAGAAGAAA TCAACAACCT GAAGAACAAA ATTATAGTGA ACAACGTGGG	1620
45	AATTCTAAAA TATCAGCA AAGTATAAAA TATAAAGATC ATTCACATTA CCATACGAAT	1680
	AAGCCAGGTA CATATGTTTC TGCAATTAAT GGTATTGAGA AGGAAACGCA CAAGCCAAAA	1740
50	ACACATAATA TGTATTCTAA TAATACAAAT CATCGTGCTA AAGATTCAAC TCCAGATTAT	1800
	CACAAAGAAA GTTTCAAGAC TTCAGAGGTA CCGTCAGCTA TTTTGGCAC AATGAAACCT	1860

55



EP 0 786 519 A2

	AAACAAAAAT ATGATAAATA TGTAGCTAAG ACGCAAACGT CTCAAAATAA ACAATTAGAA	1980
	CAAGAAAAAC AAAATGATAG TGTGTGCTAAA CAAGGAACTG CATCTAAATC ATCTGATGAA	2040
5	AATGTATCAT CAACAACAAA ATCAATGCCT AATTATTCAA AAGTTGATAA TACTATCAAA	2100
	ATTGAAAATA TTTATGCTTC ACAAATTGTT GAAGAAATTA GACGTGAACG AGAACGTAAA	2160
	GTGCTTCAAA AGCGTCGATT TAAAAAGCG TTGCAACAAA AGCGTGAAGA ACATAAAAAAC	2220
10	GAAGAGCAAG ATGCAATACA ACGTGCAATT GATGAAATGT ATGCTAAACA AGcGGAACgC	2280
	TATGTTGGTG ATAGTTCATT AAATGATGAT AGTGACTTAA CAGATAATAG TACAGATGCT	2340
	AGTCAGCTTC ATACAAATGG CATAGAGAAT GAACTGTAT CAAATGATGA AAATAAACAA	2400
15	GCGTCAATAC AAAATGAAGA CACTAATGAC ACTCATGTAG ATGAAAGTCC ATACAATTAT	2460
	GAGGAAGTTA GTTGAaTCA AGTATCGACA ACAAACAAT TGTCAGATGA TGAAGTTACG	2520
	GTTTCGAATG TAACGTCTCA ACATCAATCA GCACTACAAC ATAACGTTGA AGTAAATGAT	2580
20	AAAGATGAAC TAAAAATCA ATCCAGATTA ATTGCTGATT CAGAAGAAGA TGGAGCAACG	2640
	aATAAAGAAG AATATTCAGk AAGTCAAATC GATGATGCAG AATTTTATGA ATTAAATGAT	2700
25	ACAGAAGTAG ATGAGGATAC TACTTCAAAT ATCGAAGATA ATACCAATAG AAACGCGTCT	2760
	GAAATGCATG TAGACGCTCC TAAAACGCAA GAGTACGCAG TAACTGAATC TCAAGTAAAT	2820
	AATATCGATA AAACGGTTGA TAATGAAATT GAATTAGCAC CGCGTCATAA AAAAGATGAC	2880
30	CAACAAACT TAAGTGTCAA CTCATTGAAA ACGAATGATG TGAATGATAA TCATGTTGTG	2940
	GAAGATTCAA GCATGAATGA AATAGAAAAG AATAACGCAG AAATTACAGA AAATGTGCAA	3000
	AACGAAGCAG CTGAAAGTGA ACAAATGTC GAAGAGAAAA CTATTGAAAA CGTAAATCCA	3060
35	AAGAAACAGA CTGAAAAGGT TTCAACTTTA AGTAAAAGAC CATTTAATGT TGTCATGACG	3120
	CCATCTGATA AAAAGCGTAT GATGGATCGT AAAAAGCATT CAAAAGTCAA TGTGCCTGAA	3180
	TTAAAGCCTG TACAAAGTAA GCAAGCTGTG AGTGAAAGAA TGCCTGCGAG TCAAGCCACA	3240
40	CCATCATCAA GATCTGATTC ACAAGAGTCA AATACAAATG CATATAAAC AAATAATATG	3300
	ACATCAAACA ATGTTGaGAA CAATCAACTT ATTGGTCATG CAGAAACAGA AAATGATTAT	3360
	CAAAATGCAC AACAATATTC AGAGCAGAAA CCTTCTGTTG aTTCAACTCA AACGGAAATA	3420
45	TTTGAAGAAA GTCAAGATGA TAATCAATTG GAAAATGAGC AAGTTGATCA ATCAACTTCG	3480
	TCTTCAGTTT CAGAAGTAAG CGACATAACT GAAGAAAGCG AAGAAACAAC ACATCCAAAC	3540
50	AATACTAGTG GACAACAAGA TAATGATGAT CAACAAAAG ATTTACAGTC ATCATTTTCA	3600
	AATAAAAAATG AAGATACAGC TAATGAAAAT AGACCTCGGA CGAACCAACA AGATGTTGCA	3660

55

CCAAGTGTTC CATTACTAGA AGAACCACAA GTTATTGAGT CGGACGAGGA CTGGATTACA 3780  
 GATAAAAAGA AAGAACTGAA TGACGCATTA TTTTACTTTA ATGTACCTGC AGAAGTACAA 3840  
 5 GATGTAAGTG AAGGTCCAAG TGTTACAAGA TTTGAATTAT CAGTTGAAAA AGGTGTTAAA 3900  
 GTTTCAGAAA TTACGGCATT ACAAGATGAC ATTAAAATGG CATTGGCAGC GAAAGATATT 3960  
 CGTATAGAAG CGCCTATTCC AGGAACTAGT CGTGTGGTA TTGAAGTTCC GAACCAAAAT 4020  
 10 CCAACGACAG TCAACTTACG TTCTATTATT GAATCTCCaA GTTTTAAAAA TGCTGAATCT 4080  
 AAATTAACAG TTGCGATGGG GTATAGAATT AATAATGAAC CATTACTTAT GGATATTGCT 4140  
 AAAACGCCAC ACGCACTAAT TGCAGGTGCA ACTGGATCAG GGAAATCAGT TTGTATCAAT 4200  
 15 AGTATTTTGA TGTCTTTACT ATATAAAAAT CATCCTGAGG AATTAAGATT ATTACTTATC 4260  
 GATCCAAAAA TGGTTGAATT AGCTCCTTAT AATGGTTTGC CACATTTAGT TGCACCGGTA 4320  
 ATTACAGATG TCAAAGCAGC TACACAGAGT TTAAATGGG CCGTAGAACA AATGGAACGA 4380  
 20 CGTTATAAGT TATTTGCACA TTACCCATGT ACGTantATA ACAGCATTTA ACnAAAAAGC 4440  
 CCCATATGAT GAAAGAATGn CAAAAATTGT CATTGTaATT GATGAGTTGG CTGATTTAAT 4500  
 25 GATGATGGTC CGCAAGAAGT TG 4522

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 751 base pairs  
 30 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

TCAAGTTTAC GGATACGTAT ATATTTTGCA TGACATTTAG TGCAATAATA TTCATAATTT 60  
 GCCCGTTGTT GATAGCTTTC AATGCTGTTA CAAAATCTAG GCGCTCCAAC CTGTTGGCTC 120  
 40 AATCGTTTAA AATCTTGATC TTTATGTTGA TAACCTTTAC CAGCAATATG CAAGTGATAA 180  
 TGACACAATT CGTGCAGTAT AATTTTACAC ACAGCATCTT CTCCATAATG CTCATATTGT 240  
 TTTGGATTAA TTTCAATATC ATGGGACTTT AAAAGATAAC GTCCGCCTGT TGTACGTAAC 300  
 45 CTTTATTAA AATATGCACA ATGTCGAAAC GTACGTCCAA ATTTTCTTC CGAAAGATTC 360  
 TCAACCATTG GCTGAAGTTT GTCATTATTC ATGTGGATCA ATCATCGTTA ATGATACTTT 420  
 50 GTCTTTATTT TTGTCAATAC TGTAAATCCA AACGTCAACG ATATCACCAA CACTGACAA 480  
 ATCCATTGGA TTTTTCACGA ACTTCTTAGA AAGTTTCGAA ACATGGACAA GTCCATCTTG 540

55

TTTCATTCCCT TCTTGTAAT CTTCAATTGA TAGCACATCG GATTTAAGGA TTGGTGTTC 660  
 AAACTCGTCC CTTGGATCTC GATTAGGTGC GTTCAAGGAT TTAATAATAT CCTCTAATGT 720  
 5 AGGTACACCG ACTTGTAATT CAATCGCCAG T 751

## (2) INFORMATION FOR SEQ ID NO: 41:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1076 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

TCTCCAGCTT TAACTTGATC TGGCACTTTA ACAATTGTCT GATCCATACA TACGCGACCA 60  
 20 ATAACCTCGC ATTGATGACC ATTTACATTT ACAAAGCTAC CTTGCATTAT GCGTAAATGG 120  
 CCATCTGCAT ATCCAATAgG TAACAATGCT ATTGTAGTTG GGTCAGTAGC TGTATAAGTT 180  
 GCACCATAAC TTACAGACTC ACCCGCTTGT AGCGTCTTTG TTTGAACTAC ATTAGCAATT 240  
 25 AATTGCACAC TTGGTTTAAG GTGTACTTTA ACTTTTGTCT GTACATACTC TGATGGATAA 300  
 TATCCATAAA GGGAAATTCC TGGTCTTATT GCATTACAGA ATTGGCAATC CATTAAATAGA 360  
 GAGCCTGCTG AGTTCTGACA ATGTATATAT TCAGGTTTAA TTGCTTCATT GACCATATCT 420  
 30 TTAAAACGTT GATATTGTTC AGTTGTCATA TCTCCTGGTT CGTCAGCACA GGCAAAGTGT 480  
 GTAAACACGC CTTCAAATAC AAGTTGCTCA TATTGTTGAA TGATTTCAAT CACTTCTTGA 540  
 TACGTTTTAG TATCTTTAAT ACCTAAACGT CCCATTCTCTG TATCTAATTT AATGTGCAAC 600  
 35 CATAACTTTT TCTCTTGCTC ACCAGAAATG TTTTAAATTG CTTCTTTCAA CCACTGTTTA 660  
 GACCGAACCG TTAAGGCAAC TCGGTGTTGT ATCGCTTTAT CAATATCTTT AGCTGGTAAC 720  
 ACACCTAAGA CTAAAATTTT AGCAGTAATC CCATGCATTC TAAGTTCTAT CGCTTCATCT 780  
 40 AACGTTGCTA CAGCAAAAAA TGTGGCGCCA TTTCCATTA AATGACGTGC TACTTTAACA 840  
 CTACCTAGTC CATAGGCATT GGCTTTAACG ACAGCCATCA CTGTTTTATT TGGATGCAAT 900  
 GTACTGAATA CTTTGAAATT TGATGCAACA GCGTTTAAAT CTACATTCAT ATACGCAGAT 960  
 45 CTATAATATT TATCCGACAT ATTACTTCCT CCTGTAATTC CCACACGTTT TAAACTAGA 1020  
 TCTTAATTAT CATTGTATAA CAAATTTAAA ATGCTGACTT TTCTAAAACA ACTTGG 1076

## (2) INFORMATION FOR SEQ ID NO: 42:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2930 base pairs

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

	TGACCACAAT GCCCAATACA ACCATCCCAT GGTAAAGCCA AGAGATGAGT CAATAAAGCG	60
	TGTTGAATAA GAGCTGAATG AACCTGATAC TGGATAAAAT GTTGCCAACCT CTCCAATTGA	120
10	TGACATTAAG AAATATAGCA TGACACCAAT AACAAGATAA GCGAGTATAG CGCCTCCAGG	180
	ACCAGCTTGA GAAATGATAT TACCAGTAGC TACAAATAGA CCAGTCCCAA TTGCACCACC	240
15	TATAGCAATC ATGGAATGT GTCTTGAGTT AAGACTACGG TTCATTTTAT TATCTTCCAT	300
	ATTTAGTCTC CCATCTATTT AAATATACCC ATTATTGTAA GCTTTTAAAG TGTACTATTC	360
	AATAACTATT TAGTACTGTA AAGCGAAAAA ATTAAATTT TCTGATTTTT TAATCATCTT	420
20	GAGCATGTTT AATTGTAATT TTGATGGGGT TAAATTATAA TATGTATTAA ATTATAATTA	480
	TnATAAATTG TGGAGGGaTG ACTATGTCAC AACAAGACAA AAAGTTAACT GGTGTTTTTG	540
	GGCATCCAGT ATCAGACCGA GAAAATAGTA TGACAGCAGG GCCTAGGGGA CCTCTTTTAA	600
25	TGCAAGATAT TTACTTTTTA GAGCAAATGT CTCAATTGTA TAGAGAAGTA ATACCAGAAC	660
	GTCGAATGCA TGCCAAAGGT TCTGGTGCAT TTGGGACATT TACTGTAACCT AAAGATATAA	720
	CAAAATATAC GAATGCTAAA AtATTCTCTG AAATAGGTAA GCAAACCGAA ATGTTTGCCC	780
30	GTTTCTCTAC TGTAGCAGGA GAACGTGGTG CTGCTGATGC GGAcGTGACA TTCGAGGATT	840
	TGCGTTAAAG TTCTACACTG AAGAAGGGAA CTGGGaTTTA GTAGGGAATA ACACACCaGT	900
	ATTCTTCTTT AGAGATCCAA AGTTATTTGT TAGTTTAAAT CGTGCGGTGA AACGAGATCC	960
35	TAGAACAAAT ATGAGAGATG CACAAAATAA CTGGGATTTT TGGaCGGGTt TCCAGAAGCA	1020
	TTGCACCAAG TAACGATCTT AATGTCAGAT AGAGGGATTC CTAAAGATTT ACGTCATATG	1080
40	CATGGGTTCG GTTCTCACAC ATACTCTATG TATAATGATT CTGGTGAACG TGTTTGGGTT	1140
	AAATTCCATT TTAGAACGCA ACAAGGTATT GAAACCTTAA CTGATGAAGA AGCTGCTGAA	1200
	ATTATAGCTA CAGATCGTGA TTCATCTCAA CGCGATTTAT TCGAAGCCAT TGAAAAAGGT	1260
45	GATTATCCAA AATGGACAAT GTATATTCAA GTAATGACTG AGGAACAAGC TAAAAACCAT	1320
	AAAGATAATC CATTTGATTT AACAAAAGTA TGGTATCAG ATGAGTATCC TCTAATTGAA	1380
	GTTGGAGAGT TTGAATTAAA TAGAAATCCA GATAATTACT TTATGGATGT TGAACAAGCT	1440
50	GCGTTTGAC CAACTAATAT TATTCCAGGA TTAGATTTTT CTCCAGACAA AATGCTGCAA	1500
	GGGCGTTTAT TCTCATATGG CGATGCGCAA AGATATCGAT TAGGAGTTAA TCATTGGCAG	1560

55

5 GGTCAAATGC GCGTAGTTGA CAATAACCAA GGTGGAGGAA CACATTATTA TCCAAATAAC 1680  
 CATGGTAAAT TTGATTCTCA ACCTGAATAT AAAAAGCCAC CATTCCCAAC TGATGGATAC 1740  
 GGCTATGAAT ATAATCAACG TCAAGATGAT GATAATTATT TTGAACAACC AGGTAAATTG 1800  
 TTTAGATTAC AATCAGAGGA CGCTAAAGAA AGAATTTTTA CAAATACAGC AAATGCAATG 1860  
 GAAGGCGTAA CGGATGATGT TAAACGACGT CATATTCGTC ATTGTTACAA AGCTGACCCA 1920  
 10 GAATATGGTA AAGGTGTTGC AAAAGCATTG GGTATTGATA TAAATTCTAT TGATCTTGAA 1980  
 ACTGAAATG ATGAAACATA CGAAAACCTT GAAAAATAAA TTTGATATGT AGTTTCTATA 2040  
 TTGCGTAGTT GAGCAGTTTA TGATATCATA ATAAATCGTA AAGATTCTTA ACAAGAGAGG 2100  
 15 GTGTTTAACG TCGCGTAAA CGTAACATTA GCATGCACAG AATGTGGCGA TCGTAACTAT 2160  
 ATCACTACTA AAAATAAACG TAATAATCCT GAGCGTATTG AAATGAAAAA ATATTGCCCA 2220  
 AGATTAAACA AATATACGTT ACATCGTGAA ACTAAGTAAT TCTTATCATT CAAATACGAC 2280  
 20 GATTTGAAAA TAAAGCGGGC TTACCTATTA TATTGGGGAG CTCGCTTTTT TATGAAATTT 2340  
 TTGTGAAGAG TGATTAATGG ATTGAGTTTC ATCGGTAGAA CAATATATGA TTATATTAGT 2400  
 TGTACTTTTA TTAATAATTG AGAATATTTA TAGAAGGAAA TAGATTACTG ATTTTATAAA 2460  
 25 GTCACCTTGT TAGCGAATGC TTGAAAGAGT ATTTAATATA GTAGAATTTA AAATTTCAAA 2520  
 GCGGAATTTA ATAAGTACGA AGTAGTTCTG GGTATGTTTT ATAAATGTTT GATAATACAC 2580  
 TTTAATCTTA AATATGATGG TTTAGAAAAT GATTTAACAA AGAAATGAaa CTTTACTGTT 2640  
 30 GAATTATGTG AGGATTGTGT TATTATATAA ATCGTAATAA TTACGATTTG ATAAAAAGTG 2700  
 AGGTAACTAT ATATGGCTAA GAAATCTAAA ATAGCAAAAG AGAGAAAAAG AGAAGAGTTA 2760  
 GTAAATAAAT ATTACGAATT ACGTAAAGAG TTAAAAGCAA AAGGTGATTA CGAAGCGTTA 2820  
 35 AGAAZATTAC CAAGAGATTC ATCACCTACA CGTTTAACTA GAAGATGTAA AGTAACTGGA 2880  
 AGACCTAGAG GTGTATTACG TAAATTGAA ATGTCTCGTA TTGCGTTTAG 2930

40 (2) INFORMATION FOR SEQ ID NO: 43:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3606 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

50 CTTCTTGCCA TGGCTCTCTT TATTTAAAAA TGCTTCCAAC TTGTCCATTT GATTGTTTCT 60

55

	TTATAAAAAA CTAATTTTAC AAATGCTTTT GCGTTCTTAC AAAAAATGCA TTTGACTATT	180
	ATTATAATAA GCGTATAATT GTCGCATATT ATTTTGTGTA TTTTGGCAA TAACGAAGGA	240
5	GTATTTATGA ATAAAGACAA GCAATTGCAC AACGACAAAA TCAATCTATC CCAATTAGTC	300
	TTATTAGGGT TAGGCTCTTT AATAGGATCT GGTGCTAT TTGGTGGTG GGAAGCATCA	360
	TCAATAGCTG GACCAGCAGC AATCATATCA TGGGTCTTG GATTCTAGT CATTGGAACC	420
10	ATTGCCTATA ACTACATTGA AATCGGCACA ATGTTTCTC AATCAGGTGG CATGAGTAAC	480
	TATGCCCAGT ATACACATGG CTCATTATTA GGCTTTATTG CTGCTTGGGC GAATTGGGTG	540
	TCCTTGGTGA CAATAATACC TATCGAAGCT GTGTCAGCTG TTCAATATAT GAGTTCTTGG	600
15	CCGTGGCATT GGGCGAAACC AATGAGATAT TTAATGGAAA ATGGCTCTAT TAGCACATAC	660
	GGATTGCTAG CTGTATATCT CATCATTTGT ATTTTTCAT TATTAACTA TTGGTCCGTA	720
	AACTTTTAA CATCATTTAC GAGTTTAATT TCTGTATTTA AATTAGGCGT ACCCATGTTA	780
20	ACCATCATCA TGTGATGCT ATCAGGATTC GACACTTCAA ATTACGGCCA TTCGCAAGC	840
	ACATTTATGC CTTACGGAAG TGCACCGATT TTTGCTGCAA CAACAGCATC AGGGATTATT	900
	TTTTCATTCA ATTCATTCCA GACAATTATT AATATGGGTT CAGAAATTAA AAATCCTGAA	960
25	AAAAATATCG CAAGAGGCAT CGCTATCTCA CTGTCAATCA GTGCAGTGT GTACATCATT	1020
	TTACAAAGTA CGTTTATCAC TTCTATGCCT CAATCAATGT TACAACATAG TGGATGGAAT	1080
	GGCATCAACT TCAATTCACC ATTTGCTGAT TTAGCTATCT TATTAGGAAT TAATTGGCTC	1140
30	GCAATTTTAC TATACATTGA AGCTTTTGTG TCACCATTCG GTACTGGCGT GTCATTTGTC	1200
	GCCGTTACAG GTCGAGTTTT ACGAGCAATG GAGAAAAATG GACATATCCC TAAATTTCTT	1260
35	GGGAAGATGA ATGAAAAATA TCATATCCCA CGTGTAGCAA TCATCTTTAA TGCCATCATT	1320
	AGTATGATTA TGGTTACATT ATTTAGAGAT TGGGGTACGC TAGCAGCAGT TATTTCTACT	1380
	GCAACTTTAG TAGCCTATTT AACTGGCCCA ACGACAGTGA TTGCATTAAG AAAAATGGGA	1440
40	CCAACAATGA CTCGTCCATT TAGAGCAAAA ATTTTAAAAG TAATGGCACC ATTATCATT	1500
	GTATTAGCTT CATTAGCTAT ATATTGGGCA ATGTGGCCAA CAACGGCTGA AGTTATTTA	1560
	ATCATTATAC TTGGATTACC AATCTACTTC TTCTATGAAT ATCGTATGAA TTGGCGTAAT	1620
45	ACAAAGAAAC AAATTGGTGG TAGCTTATGG ATTATTGTAT ATTTAATCGT GCTATCAATA	1680
	CTGTCAATTA TAGGAAGCAA AGAATTTAAA GGCTTAAATA TGATTCATA TCCATTTGAC	1740
	TTTATCGTTA TTATTATTGT GGCACTTATC TTCTATTACA TCGGTACAAC GAGTTCATT	1800
50	GAAAGCGTCT ATTTCCGTCG CGCAACACGA ATCAATACGA AGATGCGTGA GTCACTAAAT	1860

55

EP 0 786 519 A2

	CACACACATT AACCAACCAT TGATTTCAAC ATCTTGGTTG GTTTTTTATT TTGAAAATCG	1980
	GTTATAAATA ACTAACATAA CAAGATGATG ATCAGGCTGG GACATAAATC AATGTTCTAT	2040
5	GCTCTACGAA gTTATATTGG CAGTAGTTGA CTGAACGAAA ATGCGCTTGT AACAAAGCTTT	2100
	TTTCGATTCT AGTCAGGGGC CCCAACACAG AGAATTTCTGA AAAGAAATTC TACAGGCAAT	2160
	GCAAGTTGGG GTGGGACGAC GATAAAGAAA TACTTTTTCT ATAGAAATTA GTATyCTTA	2220
10	TGCATGAGTT TTA CTCTCATGT ATTCATATTT TTAAGTACAC ATTAGCTGTG GCTAATGTAT	2280
	AAGAACCACT ACATAATAAA TCATTTGTGG CTCTTTATCA TTTCTGTCCC ACTCCCGTAG	2340
	AAGTACATCA TATAATGCTG AAAATGGTTT GAGTTAAAAC AGATATCAAG CTCGTCTGAT	2400
15	TCAGTCACAA AATTGTCTTG TTATACTTGT CACCTATCAT CTATAGACCG TGGTATGATT	2460
	AAATGGGGA TGATAAAGGA GGTAAATAAA TATGAAGATT AATACTACAG GTGGTCAAAT	2520
	TCATGGTATT ACACAAGATG GTTTAGATAT CTTCTTAGGC ATTCCTTATG CAGAACCACC	2580
20	AGTTCATGAC AATCGCTTTA AACATTCTAC GTTAAAAACA CAATGGTCAG AGCCAATTGA	2640
	TGCAACTGAA ATACAACCCA TCCCACCGCA ACCAGACAAC AAATTAGAAG ATTTTTTCTC	2700
25	CTCACAATCT ACAACTTTTA CTGAACATGA AGACTGTTTA TATCTAAATA TTTGGAAACA	2760
	ACATAATGAT CAGACGAAGA AACCTGTCAT CATTTATTTT TATGGTGGTA GTTTTGAAAA	2820
	TGGTCATGGT ACAGCCGAAC TCTATCAACC GGCACATTTA GTACAAAATA ACGACATTAT	2880
30	CGTTATTACA TGCAATTATC GTTTAGGCGC ATTAGGATAT TTAGACTGGT CATATTTTAA	2940
	TAAAGATTTT CATTCCAATA ATGGCCTTTC AGATCAAATC AATGTCATAA AATGGGTGCA	3000
	TCAATTTATT GAATCCTTCG GTGGCGACGC TAATAACATT ACTTTAATGG GTCAGTCTGC	3060
35	AGGCAGTATG AGCATTTTGA CTTTACTTAA AATACCTGAC ATTGAGCCAT ACTTCCATAA	3120
	AGTCTTCTA CTAAGTGGCG CACTACGATT AGACACCCTT GAGAGTGCAC GCAATAAAGC	3180
	ACAACATTTT CAAAAAATGA TGCTCGATTA TTTAGATACA GATGATGTTA CATCATTATC	3240
40	GACAAATGAT ATTCTTATGC TGATGGCGAA gcTAAACAA TCTCGAGGAC CTTCTAAAGG	3300
	GCTTGATTTA ATATATGCGC CTATTAAAAC AGATTATATA CAAAATAATT ATCCAACAAC	3360
	GAAACCAATT TTTGCATGTT ATACAAAAGA TGAAGGCGAT ATTTATATTA CTAGTGAACA	3420
45	GAAAAAATTA TCGCCGCAAC GCTTTATCGA CATTATGGAA TTAAATGATA TTCCTTTAAA	3480
	ATACGAAGAT GTTCAGACGG CGAAGcAACA ATCTTTAGCG ATTACACATT GTTATTTCaA	3540
	ACAGCCGATG aAGCAATTTT TACmACmACT CAATATACmA GATTCCAACC GCACCAACTA	3600
50	TGGCTT	3606

55

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15109 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

10	GAAATTAAAA AAGCAATTGG nACAAGATGC AACAGTGTCA TTGTTTGATG AATTTGATAA	60
	AAAATTATAC ACTTACGGCG ATAAGTGGG TCGTGGTGGG GAAGTATTAT ATCAAGCATT	120
	TGTTTGTAAA ATGCAACsAG AACAAACAAA GTTAACTGCA AAAGCAGGTT GGGCTGAAGT	180
15	GAAACAAGAA GAAATTGAAA AATATGCTGG TGATTACATT GTGAGTACAA GTGAAGGTAA	240
	ACCTACACCA GGATACGAAT CAACAAACAT GTGGaAGAAT TTGAAAGCTA CTAAAGAAGG	300
	ACATATTGTT AAAGTTGATG CTGGTACATA CTGGTACAAC GATCCTTATA CATTAGATTT	360
20	CATGCGTAAA GATTTAAAAG AmAAATTAAT TAAAGCTGCA AAATAATTCA GCTATATAAG	420
	TTAGTGAAAT GAGAGTCTGA AACATATCAA TCTTTTGATA TTGTATTAGG CTCTTATTTT	480
	TATAGCTAGA AAGTTAGATA TTTGTATTTT TTTAAATAAT AAGTGCCGTT GTTATCGTTC	540
25	AATTTAATTA ATGATAGATT AGTATTATTA TAGCTAAAGT AGTATACCTG AGAAAATAGC	600
	TCAATGTATC TCTTTATTAA TAAGTTATAT CATAATTATT TTAGTGATA CTTTATGGAA	660
	GGGATATCAG GGAATGGCTT TCAATTAAAG AAGAGGTTTA AAAGGATTAC AACAGAATGT	720
30	TATGATTTTG TAGAAAGATA TATAACAACG TTTTATAAAA ACATAATATT GTTAATGGAA	780
	AATGAAATGT AAGGGGGATT TCGAGTGAAT AAGAAAGTTT ATTTTAACCA CGATGGTGGT	840
35	GTAGATGATT TAGTATCTCT ATTTTATTATTA TTACAAATGG AAAACGTTCA ATTGATAGGG	900
	GTCAGTACAA TTGGTGCTGA TTGTTATTTA GAGCCATCTT TGAGCGCATC AGTAAAAATT	960
	ATTAAATCGTT TTTCAAATGA AGATATTCAA GTTGCGCCAT CATATGAACG AGGAAAAAAT	1020
40	CCATTTCTTA AAGAATGGCG TATGCATGCC TTTTATATGG ACGCATTGCC AATTTTAAAT	1080
	GAGCCAGTCA AACATGTTGC TTCAAATGTG AGCGACAAAG AAGCCTTTGA AGACATTATT	1140
	CAAACTTTAA AGAGACAATC AGAAAAAGTA ACATTATTAT TTACAGGCCC GCTTACAGAT	1200
45	TTAGCAAAAG CACTACAAAA AGATTCATCT ATCGTTCAGT ATATAGAAAA ATTAGTTTGG	1260
	ATGGGTGGCA CCTTTTACC AAAAGGAAAT GTTGAAGAAC CTGAGCATGA TGGTTCTGCA	1320
	GAATGGAATG CATATTGGGA TCCAGAAGCG GTTAAATTTG TTTTGTATAG CGATATAGAG	1380
50	ATTGATATGG TTGCTTTAGA AAGTACGAAT CAAGTACCGC TAACGTTAGA TGTTAGACAA	1440

55



## EP 0 786 519 A2

	GTACCACCAT TAACACACTT TATAACAAAT TCTACTTACT TTTTATGGGA TGTTTTAACG	1560
	ACTGCTTATA TTGGTAACAA GGACTTGGTT CATTCAATTG AGAAAAAGT CGATGTAATA	1620
5	AGTTATGGAC CAAGTCAAGG TAAGACATTT GAGTGTAAG ATGGGCGCAA AATTAATGTC	1680
	ATAAATCATG TAGATAACAA CGCATTTTTT GATTATATAA CTGCACTTGC TAAAAAGTA	1740
	AATTAACAGC TGTGTAGAAT AATTAAGGTT TTAATTTATA TAGAACAACT TATTGTAAAC	1800
10	TTTTCATTTT TAAAGTTTA CAATGGTGCT ATAATAATGG TCATGAAATA CGAAAGGAAG	1860
	TAAAAATGA CAACAAAACA GTTAGTATAT ACAGCTTTAA TGACAGCGAT TATCGCTATT	1920
	TTAGGATTGG TACCGGTAAT TCCACTACCA TTTTCTTCAG TACCAATTGT ACTTCAAAAC	1980
15	ATTGGTATTT TCTTAGCAGG TGCATTTTTA GGACGTAAAT ATGGCACATT AAGTGTTATC	2040
	GTCTTTTTAT TATTAGTAGT TGCTGGCTTG CCATTGTTAT CAGGTGGTCG CGGTGGCATC	2100
	GGTGTATTCG CAGGTCCTTC AGCAGGGTTT TTAATTATAT ATCCAGTTGT AGCATTTCATG	2160
20	ATTGGGGCGA TTCGAGATAG ATTCATCAAT GAAATTAATT TCTGGATTTT ATTCGTTGGT	2220
	ATTTTAGTTT TTGGTGTTAT AGCATTAGAT GTTATTGGTA CATTGATTAT GGGCATGATT	2280
	ATTAACATAC CATTTACGAA AGCTATTTCA ATTTCAATTAG CTTATTTGCC TGGTGATATA	2340
25	TTAAAAGCAA TTGTAGCAAG TTTGATTGGT ACAGCTTTAC TTAATCACTC GCAGTTTCGT	2400
	CAAATTATGG GAATAAAATA ATCATATTTA AGATAGTAAA GTAATTGAAT AAGTTGCTTT	2460
30	GAAATTTATA AAAGTGAAAG GAGTAGGTGT CAATGGCTAG TATAAGTATG TCAGATATAT	2520
	ATTGTAACGG CACTATATTT GAAAATGACG ACGAGCAGTT GATTTATTTA ACGCCTTCTT	2580
	TTCCACAACG ATACACAAGT AACACATGGA TATATAAAAA GACGCCTACC CAAGAGCGAT	2640
35	GGCTGAAAGA CTTAGAACGT CAACATCAAT TACATACAAA TCAAGGTTCA AATCATTATG	2700
	CGTTTAGTTT CCCGAAAAAT GAACAACTTG ATAATCATTG GATGGCTATG TTAAAGATA	2760
	TGAATTTTGA ACTAGGTATT ATGGAATTGT ATGCCATAGA AAGTGATGCG CTTGCCAATT	2820
40	TGCCGCGTAA CTCTGACGTT GAAATTGCCA TCGTTGACGA GTCGCATATA GATGCCTATT	2880
	TAAAAGTTGC ATATCAGTTT AGTTTGCCAT TTGGAAAAGA CTATGCAGAT GCACATGAAG	2940
	AAATGGTAAG GGAACATTAT CAAAAAGATG TGATTAAACG CTTAGTAGCT TATTTAAATA	3000
45	ATGAACCTAT TGGCGTTGTA GATGTCATTG AAAGTGAAAA TTACATTGAA TTAGATGGAT	3060
	TTGGTGATT AGAACAATTT CGGCACCAAG GAATTGGATC TACAATTCAA TCGTTGATAG	3120
	GTGAATACGC CATATCAAAA AATCACAAAC CAATCATATT AGTTGCAGAT GGTGAAGATA	3180
50	CAGCAAAAGA TATGTATGCA AAGCAAGGTT ATGTCTATCA ATCGTTTTGT TATCAAATAT	3240

55

EP 0 786 519 A2

	TAAGCTGGTT TCGAGTAGAA ATCAACTTAC TGCTTTTAA ATTGTTTGA GCTACTTATA	3360
	CTTATAAAAA TAGTGC GTTT AAATTGTTGA TTCATGTAGA ATATCGTTCA TTATGACACA	3420
5	CTATAATGAA TATGTTATTG TTCAGAATCA ATGATACGTT CTGGATGACT GTATATATTA	3480
	AAGCCACCAT TTCGAATAAA TCCAAC TGCC GTAATATTTA GGTCAATTAGC TAAGGTTACA	3540
10	GCAAGCGTTG TCGGAGCTGA TTTAGATAAA ATGACGCCAA CACCAATTTT TGCGGCTTTA	3600
	ATTAAATTT CTGATGAAAT ACGTCCACTA AAAATTAATA CTTTATCTCG GACAGTAATA	3660
	TGTCGCTGAA TACAAAATCC ATATAATTTA TCTAGAGCGT TATGTCTACC AATGTCTTGT	3720
15	CGATGTACAA AAAATGTCAA ACCATCGCTT ATAGCAGCAT TATGTAAGCC ACCTGTTTCT	3780
	TGGTAAATAT GACTTGCACT TTGTAATCGA GTCATCATGT TAATAATTTG CATTGGAGTT	3840
	AAAGTGATTT TAGACATAGA TGTTTTAGCG ATAGCAGCAT CATTTTGAAA ATAAAACTCA	3900
20	CGACTCTTTC CGCAACAAGA TGCAATCATT CGTTTTGTGG AATATTGAAA GCGATCGCCT	3960
	AAATCTTTAT TAAGTTCAAC ATGGGCAAAA CCTTTACTAT CATCAATCAG TACAGATTTT	4020
	AATTCATCTC GCTTTAAAAT GGCACCTTCC GAAGCCAGAA ATCCAATGAC TAACTCCTCA	4080
25	AGGTTTGTG GACTGCATAT AACAGTCGCA AATTCTTCAC CATTCACCAT AATTGTAAGT	4140
	GGAAATCTG TCACATATTG ATCTGTTGTA TTGAATAATT TTCCATCTTC ATATCTAACA	4200
	ATTGGTTGAC CTAAAGATAC ATCTTTGTTC ATTATCTAAC CCCTTTAATT AGCTTAACT	4260
30	TTATTTTAAA GCAATTTGCT TAAAATTTA ACATATTGTC TTAAGTTTGA AATTTGATTG	4320
	ATAAAAATTA ATAGCGAGCA ATCTGTTTGA TTTAAATTGA ATTCGAGAAT ATACATACTA	4380
35	GGGCATCAAT TAATAAATAT CAATCTTATG CAAATTGAC AATTGTTTGA ATCAATATAT	4440
	AAACAGGCAA CGGTTCTTTT CAAATATAAT AGTAAGTGTA TAATGAAAAT GTAAATATTA	4500
	TTAAATATGG GGGTTCCTC AATGAAATTG AAACGTTTAT TTGCTGTTGT GATTGCAATG	4560
40	CTTTTAGTAT TAGCTGGTTG CTCTAATTCT AACGATAATA ATGAAAGTAA AAAAGATGAC	4620
	GCAGACAATG GTAAGAAACA AGAGATTCAA GTTGCGAGCG CAGCAAGTTT AACAGATGTA	4680
	ACCAAGAAAT TAGCTTCAGA ATTTAAAAA GAGCATAAAA ATGCTGATAT TAAATTTAAC	4740
45	TATGGTGGAT CAGGGGCATT AAGAAAACAA ATTGAATCAG GCGCACCTGT TGACGTATTT	4800
	ATGTCTGCAA ATACTAAAGA TGTAGATGCA TTAAGACAA AGAATAAAGC GCATGATACA	4860
	TATAAATATG CGAAAAATAG TCTAGTATTA ATTGGTGATA AAGATTCAAA TTACACTTCA	4920
50	GTAAAAGACT TAAAAGACAA TGATAAATTA GCATTAGGTG AAGTGAAAAC TGTACCAGCA	4980
	GGAAAATATG CGAAACAGTA TTTAGATAAC AATAACTTAT TTAAAGAAGT CGAAAGTAAA	5040
55		

EP 0 786 519 A2

CAAGGTTTTG TGTATAAAAC TGACTTATAT AAACAAAATA AAAAAATTGA TACTGTAAAA 5160  
GTAATTAAAG AAGTAGAACT TAAGAAGCCA ATCACATACG AAGCTGGTGC TACATCAGAT 5220  
5 AGTAAATTAG CAAAAGAGTG GATGGAATTC TTAAATCAG ATAAAGCTAA AGAAATACTA 5280  
AAAGAATACC ACTTTGCAGC ATAAGGAGTT GTAATCCATG CCTGACTTAA CACCTTTTTG 5340  
GATATCAATA CGAGTTGCTG TAATCAGTAC GATTATTGTA ACGGTTTTAG GTATTTTTAT 5400  
10 ATCTAAATGG TTGTATCGTC GTAAGGGTTC GTGGGTTAAA GTATTGAAA GTTTATTGAT 5460  
ATTACCTATT GTTTGCCGC CAACGGTATT AGGTTTTATT CTATTAATCA TCTTCTCGCC 5520  
AAGAGGACCA ATCGGTCAAT TCTTTGCGAA TGTACTACAT TTACCTGTAG TGTTCACTTT 5580  
15 GACAGGTGCT GTGATAGCAT CTGTCAATTGT TAGTTTTCCA CTAATGTATC AACATACTGT 5640  
GCAAGGCTTC AGAGGTATAG ACACGAAAAT GATTAATACA GCTAGAACGA TGGGAGCAAG 5700  
20 TGAAACGAAA ATTTTCCTCA AATTAATTTT ACCATTAGCT AAACGCTCTA TTTTAGCAGG 5760  
TATAATGATG AGTTTTGCTC GTGCATTAGG TGAGTTTGGT GCTACATTAA TGGTTCAGG 5820  
ATATATTCCA AATAAACGA ATACACTACC TTTAGAAATA TACTTCTTAG TGAACAAGG 5880  
25 TAGAGAAAAT GAAGCGTGGT TATGGGTATT AGTGCTAGTC GCATTCTCTA TTGTGGTTAT 5940  
ATCTACAATT AATTTATTGA ATAAAGATAA ATATAAGGAG GTCGACTAGA TGCTTAAAT 6000  
CAATGTGAAA TATCAATTAA AGAACACTTT AATTCGCATC AATATAGATG ATACTGAACC 6060  
30 AAAAAATTTAT GCAGTTCGTG GTCCATCTGG CATTGGTAAA ACTACTGTTT TAAATATGAT 6120  
TGCCGATTA CGTAAAGCAG ATGAAGCTAT TATCGAAGTG AATGGGCAAT TACTTACTGA 6180  
TACGGCAAAA AACGTGAATG TTAAAITCA ACAACGACGT ATTGGATATC TGTTTCAAGA 6240  
35 CTACCAATTG TTTCTAATA TGACGGTCTA TAAAAATATT ACTTTTATGG CTGAACCATC 6300  
TGAAACACATC GATCAATTAA TTCAAACCTT AACATTGAT CATTTGATGA AACAATATCC 6360  
40 TATGACATTG TCAGGTGGAG AGGCACAACG TGTAGCACTT GCACGTGCAC TTAGCACrAA 6420  
ACCAGATTTA ATTTTATTAG ATGAACCTTT TTCTAGTTTG GATGATACTA CAAAAGATGA 6480  
GAGTATTACA TTAGTTAAAC GTATTTTCAA CGAATGGCAA ATACCAATCA TATTTGTGAC 6540  
45 ACATTCAAAC TATGAAGCAG AACAAATGGC TCATGAAATT ATTACAATTG GGTAATCATT 6600  
TATTTGCCAT TAAAGAGTTT AGAACGTATT TAAATTTGTA GAAGTGAATG CTTCTATCAG 6660  
CATTTTAATG ATGTTTTAAA CTCTTTTTTA GGGGCAGTTT TTTTGAGAGA CATTGACGCG 6720  
50 CGTCATATAA TGAAAGTAAT GATAAAAAGA AAGGATAACT TAATGTGAGT CAAGAACGTT 6780  
ATTCAAGGCA AATTTTATTT AAACAAATAG GTGAAATAGG TCAAAGCAAA ATAAATCAAA 6840  
55

	GAGCAGGCAT	TGCCAAACTA	ATCATTGTTG	ATAGAGATTA	TATTGAATTT	AGTAATTTAC	6960
	AAAGACAAAC	ATTGTTTACT	GAAGAAGATG	CTTTGAAAAT	GATGCCTAAG	GTGGTTGCAG	7020
5	CTAAAAAGCA	TTTGCTAGCG	TTACGTAGTG	ATGTTGATAT	TGATGATTAT	ATTGCCCATG	7080
	TGGATTATTA	TTTTTTGGAA	ACACATGGAC	AGGACGTTGA	CGTTATTATT	GATGCAACCG	7140
	ATAACTTTGA	AACACGACAA	CTGATTAATG	ATTTTGCATA	TAAATATCGT	ATACCTTGGA	7200
10	TTTATGGTGG	TGTTGTACAG	AGTACATATA	CAGAAGCTGC	ATTTATACCT	GGTAAACAC	7260
	CTTGCTTTAA	CTGTTTGGTA	CCACAATTGC	CAGCATTAAA	TTTAACATGT	GATACAGTAG	7320
15	GGGTCAATCA	ACCTGCCGTG	ACGATGGCAA	CAAGTTTACA	ATTAAGAGAT	GCGATGAAAG	7380
	TATTAACGGA	ACAACCAATT	GACACAAAAA	TAACTTATGG	CGATATTTGG	GAAGGTAGTC	7440
	ATTATTCATT	TGGTTTCAGT	AAAATGCAAC	GTTTCAGACTG	TACAACTTGT	GGAGATGTAC	7500
20	CAAGTTATCC	GTATTTAAAC	AAGAATGAAC	AACGTTATGC	AACATTGTGT	GGTAGAGACA	7560
	CTGTACAGTA	TGAAAATGCA	TCAATTACAC	ACGACATTCT	TGTTCAATTT	TTAAAAACAAC	7620
	ATCAGTTAAA	TTATCGCAGT	AATTCGTATA	TGGTTATGTT	TGAATTTAAA	GGACACCGCA	7680
25	TTGTTGCTTT	TAAAGGTGGA	AGGTTTTTAA	TACATGGCAT	GACACGCACA	TCAGATGCCA	7740
	CACATCTAAT	GAATTTATTG	TTTGGATAAA	AAAAGATAAG	ACAAAAGGAG	TGTAATATTA	7800
	TGGGCGAACA	TCAAAACGTT	AAATTGAATC	GTACAGTTAA	AGCAGCCGTA	CTAACGGTAT	7860
30	CAGATACTAG	AGACTTTGAT	ACAGATAAAG	GTGGTCAATG	CGTGCGCCAA	CTATTACAAG	7920
	CAGATGACGT	TGAAGTGAGT	GACGCACATT	ATACAATTGT	GAAAGATGAA	AAAGTAGCCA	7980
35	TCACGACGCA	GGTGAAGAAG	TGGTTAGAAG	AAGATATTGA	TGTCATCATT	ACGACTGGTG	8040
	GAACAGGTAT	TGCACAACGT	GATGTGACGA	TTGAAGCAGT	AAAACCACTT	TTAACTAAAG	8100
	AGATAGAAGG	CTTTGGGGAA	TTGTTTAGAT	ATTTGAGTTA	TGTTGAAGAT	GTTGGCACGC	8160
40	GTGCATTATT	GTCTCGTGCT	GTAGCAGGTA	CAGTTAATAA	TAAATTGATA	TTTTCGATTG	8220
	CAGGATCAAC	AGGCGCAGTT	AAATTAGCAT	TAGAAAAGCT	CATTAAACCA	GAATTAAATC	8280
	ATCTGATTCA	TGAGCTTACA	AAATAATTTA	TTGATTTGAT	TGGCGTTGAA	AATCTCCAGA	8340
45	TTTACCGCCA	GACTTGCTTT	CAAGGTAGGT	TTCGCCAATA	ATCATACCTT	TATCAACTGC	8400
	TTTCGTCATG	TCGTAAATGG	TTAAAGCCGT	TGCTGATGCA	GCGGTTAAAG	CTTCCATTTG	8460
	AACACCGGTT	TTGCCAGTTG	TAGAGACAGT	TGTTTGAATG	TTTAAAGTAT	AAAGGGGTGC	8520
50	ATTTGTTTCA	TCCCAGCTGA	AGTGAACATC	TATGCCAGTC	AATGGTAATG	GATGGCACAT	8580
	CGGAATAAGT	GTTGATGTAT	TTTTGGCAGC	CATAATACCA	GCGATTTGAG	CAGTGTTCAA	8640
55							

EP 0 786 519 A2

	AATGCTTGAA TGAGCGACAG CAGTTCTTTT TGTAATTGT TTGTCTGATA CATCGACCAT	8760
	TTTGCGGTGG CCTTGTTGAT TAATATGAGT AAACCTCAGTC ATTTTACCCC TCCTAGTGCA	8820
5	TCTAGTATAT CATGAAAAAA TAAAAGTTTT GGAGATGATT TTTAATGGTA GTAGAAAAAA	8880
	GAAACCCAAT CCCAGTTAAA GAAGCAATTC AACGTATCGT TAATCAGCAG AGTTCAATGC	8940
	CGGCAATTAC GGTAGCACTT GAAAAAAGTC TAAATCATAT CTTAGCAGAA GATATTGTAG	9000
10	CTACTTATGA TATACCAAGG TTTGATAAAT CACCTTATGA TGGTTTTGCA ATTTCGCAGTG	9060
	TTGATTCAAC AGGGGCAAGT GGTCAAGATC GCATTGAGTT TAAAGTGATT GATCATATTG	9120
	GTGCAGGTTT AGTTTCTGAT AAATTAGTTG GGGATCACGA AGCGGTGCGT ATTATGACTG	9180
15	GAGCACAAAT ACCTAATGGC GCAGATGCTG TTGTTATGTT TGAACAAACG ATTGAACTAG	9240
	AAGATACATT TACAATTCGT AAACCATTTT CAAAAAATGA AAATATATCT TTAAAAGGTG	9300
	AAGAAACAAA GACAGGCGAT GTTGTCTTAA AAAAAGGACA AGTAATTAAT CCAGGGGCTA	9360
20	TCGCGGTCCT TGCAACATAT GGCTATGCAG AGGTTAAAGT TATTAAGCAA CCGAGTGTCTG	9420
	CTGTTATTGC AACAGGAAGC GAATTATTAG ATGTTAATGA TGTATTAGAA GATGGGAAAA	9480
25	TTGCTAACTC TAATGGCCCA ATGATTCGTG CCTTAGCAGA AAAATTAGGT CTTGAAGTTG	9540
	GTATTTACAA AACACAAAAA GATGATTTAG ATAGTGGCAT CCAAGTCGTT AAAGAAGCTA	9600
	TGGA AAAACA TGATATCGTT ATTACAACGG GCGGAGTTTC TGTGGAGAT TTTGACTATT	9660
30	TACCTGAGAT TTATAAGGCT GTAAAGGCGG AAGTGTTATT TAATAAAGTA GCAATGCGTC	9720
	CTGGTAGCGT AACAAACGGT GCATTTGTAG ATGGAaAGTA TTTGTTGGa TTATCTGGAA	9780
	ATCCATCAGC TTGTTTTACA GGATTTGAAC TATTTGTGAA nCCAGCTGTT AAACATATGT	9840
35	GTGGCGCACT AGAAGTCTTC CCGCAAATAA TTAAAGCAAC ATTAATGGAA GATTTTACCA	9900
	AGGCAAAACC ATTACACGA TTTATACGTG CTAAAGCAAC GTTAACAAGT GCTGGAGCTA	9960
	CTGTAGTACC TTCAGGATTC AATAAATCAG GTGCGGTTGT AGCGATTGCA CATGCTAACT	10020
40	GTATGGTCAT GTTACCAGGA GGGTCACGTG GTTTTAAAGC GGGGCATACA GTAGATATTA	10080
	TATTGACTGA ATCTGACGCT GCTGAAGAGG AACTTCTTTT ATGATTTTAC AAATTGTAGG	10140
45	TTACAAAAAG TCTGGTAAGA CAACATTGAT GAGGCATATT GTCTCTTTCT TAAAGTCACA	10200
	TGGTTATACA GTTGCTACTA TTAAACATCA TGGGCATGGT AAGGAAGATA TTCAATTACA	10260
	GGATTGAGAC GTCGATCACA TGAAGCATTT TGAAGCGGGG GCAGATCAAA GTATTGTACA	10320
50	AGGTTTTCAT TATCAGCAAA CTGTAACACG TGTAGATAAT CAAAATCTTA CTCAAATTAT	10380
	TGAAAAATCT GTTACAATTG ACACCAATAT CGTATTAGTT GAAGGCTTTA AAAATGCTGA	10440
55		

	GAATGTTTGT TATAGCATT	ATGTAAGGGA GCATGAAGAT	TTTACAGCAT TTGAGCAATG	10560
	GTTATTAAAT AAAATTAAAA	ATGATTGTGA TACACAATTA	ACATAGAGGA TTGAAATGAA	10620
5	TGAAACAATT TGAAATCGTG	ACAGAACCGA TACAAACAGA	ACAATATCGT GAATTCACTA	10680
	TAAATGAATA TCAAGGTGCA	GTAGTTGTTT TTACCGGTCA	TGTTCCGCGAA TGGACTAAAG	10740
	GCGTCAAAAC GGAATATTTA	GAATATGAAG CGTATATTCC	AATGGCTGAA AAGAAATTGG	10800
10	CACAAATTGG AGATGAAATA	AATGAAAAAT GGCCTGGAAC	GATAACGAGT ATTGTTTCATA	10860
	GAATAGGGCC ATTACAAATT	TCAGATATCG CTGTATTAAT	TGCGGTTTCT TCACCGCATC	10920
	GTAAAGATGC CTATCGAGCA	AATGAATATG CAATTGAGCG	TATAAAAGAA ATTGTTCCGA	10980
15	TTTGAAAAA AGAAATTTGG	GAAGATGGTT CAAAATGGCA	AGGGCATCAA AAAGGGAATT	11040
	ATGAAGAAGC AAAGAGGGAG	GAATAAGAGA GATGAAGGTA	CTTTACTTCG CAGAAATTAA	11100
	AGATATATTA CAAAAGCAC	AGGAAGATAT TGTGCTTGAA	CAAGCATTGA CTGTACAACA	11160
20	ATTTGAAGAT TTATTGTTTG	AACGTTATCC GCAAATCAAT	AATAAAAAGT TTCAAGTTGC	11220
	TGTAAATGAG GAATTTGTAC	AAAAATCGGA TTTCAATCAA	CCTAATGATA CTGTTGCATT	11280
25	AATTCCACCG GTTAGTGGAG	GTTAAGGGAG CATGAAAGCA	ATAATTCTTG CAGGTGGTCA	11340
	TTCAGTGCGA TTTGGTAAGC	CCAAAGCTTT TGCGGAAGTG	AACGGTGAGA CCTTTTATAG	11400
	TAGAGTAATT AAGACATTAG	AATCAACAAA TATGTTCAAT	GAAATTATTA TTAGTACAAA	11460
30	TGCGCAATTG GCAACGCAAT	TTAAATATCC AAATGTTGTT	ATAGATGATG AGAATCATAA	11520
	TGATAAAGGT CCATTAGCAG	GAATTTATAC AATCATGAAG	CAACATCCTG AAGAAGAATT	11580
	GTTTTTTGTC GTTTCTGTTG	ATACACCAAT GATTACTGGT	AAAGCTGTAA GCACGTTGTA	11640
35	TCAGTTTTTA GTTTCTCATC	TTATTGAAAA TCATTTAGAT	GTCGCAGCTT TTAAAGAAGA	11700
	TGGACGTTTT ATTCCAACAA	TTGCATTTTA TAGTCCGAAT	GCATTAGGCG CTATAACTAA	11760
	AGCACTACAT TCTGATAATT	ACAGTTTTAA AAATGTATAT	CATGAATTAT CAACGGATTA	11820
40	TTTGGAATGTA AGGGATGTAG	ATGCGCCCTC ATATTGGTAC	AAAAATATAA ATTATCAGCA	11880
	TGATTTGGAC GCTTTAATTC	AAAAATTGTA AGCTGTTAGG	AGGTCCACAA ATGGTAGAAC	11940
	AAATAAAAGA TAAACTAGGA	CGTCCCATCC GTGACTTACG	GTTATCTGTG ACAGATCGGT	12000
45	GTAACCTTTAG GTGTGATTAT	TGCATGCCTA AAGAGGTATT	TGGAGATGAT TTCGTATTTT	12060
	TACCTAAAAA TGAACCTTTA	ACGTTTGATG AAATGGCTAG	AATCGCTAAG GTATATGCAG	12120
50	AATTAGGTGT AAAAAAATA	CGCATTACAG GTGAGAAACC	ATTGATGCGA CGGGATTTAG	12180
	ATGTACTTAT AGCTAAATTA	AATCAAATCG ATGGTATTGA	AGATATTGGT TTGACTACAA	12240

55

## EP 0 786 519 A2

	ATGTCAGTTT GGATGCTATT GATGATACGC TATTTCAATC AATCAATAAT CGTAATATTA	12360
	AAGCGACTAC GATTTTAGAA CAAATTGATT ACGCGACGTC TATTGGTTTG AATGTAAAAG	12420
5	TAAATGTTGT TATACAAAAA GGTATTAACG ATGATCAAAT CATACCAATG CTTGAATATT	12480
	TTAAAGATAA ACATATAGAG ATTGATTTA TAGAATTTAT GGATGTTGGT AATGATAATG	12540
	GATGGGATTT CAGTAAAGTT GTAACATAAG ATGAAATGCT TACAATGATA GAGCAGCACT	12600
10	TTGAAATCGA TCCTGTAGAA CCAAAATATT TTGGGGAAGT AGCAAAATAT TATCGCCATA	12660
	AGGATAATGG TGTTCATTTT GGTTTGATTA CAAGTGTTTC ACAATCATTT TGTCTACAT	12720
	GTACACGCGC AAGGCTGTCA TCAGATGGGA AGTTTTACGG ATGTTTATTT GCAACTGTCTG	12780
15	ATGGATTTAA CGTTAAAGCG TTTATTCGTT CTGGCGTGAC CGACGAAGAA TTAAAGAAG	12840
	AATTTAAAGC TTTATGGCAA ATAAGAGATG ATCGATATTC AGATGAGAGA ACTGCTCAA	12900
20	CAGTTGCCAA TCGTCAACGT AAAAAGATAA ACATGAATTA TATTGGTGGT TAATGTGTAG	12960
	GGACCACTAC ATATTAAATC ATTAGAGATG TTTTAATATT TCTGTCTTAC TCCCTAAAAT	13020
	ACAATATTAT TTATTAAAGT AAAAACGGTC ATATCTATGC CAGATTTAAT AGAAATGATC	13080
25	GTTTTTAAAG TTTTACAAG TTGGCGGGGC CCCAACACAG AAGCTGACAG AAAGTCAGCT	13140
	TACAATAATG TGCAAGTTGG CGGGGCCCCA ACATAGAGAA TTTCAAAAAG AAATTCTACA	13200
	GACAATGCAA GTTGGGGAAC GGGGCCCCAA CACAGAAGGT GACGAAAAGT CAGCATACAA	13260
30	TAATGTGCAA GTTGGCGGGG CCCCAACATA GAGAATTTCA AAAGAAATTC TACAGACAAT	13320
	GCAAGTTGGG GATCAACGAA ATAAATTTTA TGAGAATATC ATTTCTATCC CACTCTTAAG	13380
	AATCACTACA TAATAAATCT TTAGTGGTTC TTTAACATTG ATGTCACACT CCATGCCATT	13440
35	GAGTTGTAAT ATATCTTTTT TAGGTATAAA TGTTGTCGAA TAAACAACAA GTTGTCCAA	13500
	AGATAATAAT CTAAACAAGA TATAGCCAGC AATTTAATAT TTGTAATAGA TAAAATGCTA	13560
	AGTTTGATAT ATAATAAATT TAAGTAATTG TATAATAATA TGAATTACAA ACATCTAAGA	13620
40	AGAAACATAG GAGGCATCAT ATTATGAGTA ATAAAGTTCA ACGTTTTATA GAAGCAGAAA	13680
	GGGAGTTAAG TCAGTTAAAG CACTGGTTAA AAACAACACA TAAGATTTC AATTGAAGAAT	13740
45	TTGTAGTCCT TTTTAAAGTG TATGAAGCTG AAAAGATTAG CGGTAAAGAA TTGAGGGATm	13800
	CATTACATTT TGAAATGCTA TGGGATACAA GTAAAATCGA TGTGATTATC CGTAAaTCT	13860
	ATAAAAAAGA GCTTATTTCT AAATTGCGTT CTGAAACGGA TGAAAGACAA GTATTCTATT	13920
50	TCTATAGTAC TTCTCAAAAG AAATTGTTAG ATAAAATTAC TAAAGAAATA GAAGTGTTAA	13980
	GCGTTACAAA CTAAAACTT aaaaAgcaTG CCAATCTCTA TTCATCATAA TTGCGTCTTG	14040

55

GTTCATGGCA TTTCTAGTTA CATGACGTCC ATGAATTAAG AAGTAAACAA GCATAGTAAT 14160  
 GATTGCTAAA GCGGCCATAA AGCCGAAGAT TTCACTATAT GAAAACATAT GAGTAAATAA 14220  
 5 CCCAAGGAAT GATGGACCGA AGCCGACACC TGCATCTAGA CCAACGTAAA AAGTAGATGT 14280  
 CGCGATACCA TATTTAATCG GGGGTGAGAC TTTTATCGCA ATAGATTGCA TTGCAGATGA 14340  
 TAAATTTCCA TACCCTAAAC CTAGGCAAGC ACCAGCAAGT AATATTAACC AGCTTTGATA 14400  
 10 GCTTGAAATT AAGCATACAA ATGAAAGGAA AAGCATGATA AATGCTGGGT AGACAATAAT 14460  
 ATTTTCATTT TTATCATCCA TCAATCTACC AGCAATAGGT CTAGTAATTA ACGATGCTAT 14520  
 AGCATAGCAA ATAAAGAAAT AGCTTGCTGC AGTGA CTAGG TGTCGCTCTA AAGCAAATGC 14580  
 15 TTGTAAATAA GTTAGGATGG ACGCATAGGT AACGCCAATT AAAAGCATAA TTACAGCAAC 14640  
 AGGAATGGCC TCTTTTGCAA TAAATTGATG AATACTAAAT CTTGGTTTAT CAATGACATT 14700  
 AGTTTCAGTT TTGTTATTTG TTA CTTCGAA ATCAACTTTT ATAAATAATG AGATAATGAG 14760  
 20 TCCGAGTATG CCTAATATGA CACAAATAAT AAACAGTAAG TCAATTGCGT ATTTTGTAAT 14820  
 AAGTAACATG CCTAGAAATG GGCCAATCGC TGTACCTAAT ACTAAACTTA AGGAAAATAA 14880  
 ACTGATGCCT TCACTTTTTC TATTAACAGG GGTAACGTAT GCCGCAATAG TACCTGTTGC 14940  
 25 AGTTGTCACA ACTGCAGTTG CGATACCGTT TATGAGACGT ACAAAGATTA AAAAAGCTAA 15000  
 AGATCCATCA ATAAATAAAA GTAATTGCGT GATAATTAAA GCAATTAAAC CAATAAATAA 15060  
 30 TAATCGTTTA GGTCCrATTT sATTTACAAA TTTACCTGTA GCAAATCGA 15109

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9072 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GAGAGTCAAT GGCAAGAAGA ATATAAATAT TTGAGAGCGT TAATCTTTAA TGAAACAGAA 60  
 TTAGAGGAAG CGTATAAATG GATGCATCCT TGTTACACGT TGAATAATAA AAATGTAGTA 120  
 45 CTTATCCATG GCTTCAAAAA TTATGTTGCA CTATTATTTT ATAAAGGTGC CATTTTGGAG 180  
 GATAAATATC ATACACTCAT TCAACAGACT GAAAAGGTGC AAGCAGCTCG TCAGTTACGA 240  
 TTTGAAAATT TAACAGAGAT TCAAGCACGT ACCGAAGAAA TTAAATATTA TCTAGCCGAA 300  
 50 GCAATTAAAG CTGAAAAAGC TGGTAAAAAA GTTGAAATGA AGAAAAACAGA GGAATATGTT 360



EP 0 786 519 A2

	AAATTAACGC CAGGCAGACA ACATCAATAT ATATATCATA TTGGACAAGC TAAACGCagT	480
	GgAACAAAGAC AAAAGCGTGT TGAAAAGTAT ATTAACCAAA TACTAGAAGG TAAAGGGATG	540
5	CATGATAAGT AATTAATGAG TAAAGCATAC CGGTTATACA ACAACATACA AGATGACACG	600
	AAACAACCAA TGGCTCATGC TGTGGTGTGT TTTTGTAGGT GTGTCTGTCA TGGGCAACAC	660
	TTTGACGTTG GAATTCGTT ACAGGCTTGG GAGTAGAAAA TGTTAGCAA AGGCAAGGGT	720
10	GTCTACAATG AATGATGAAG ATATTAAAAT ATAAGGATGA CTTTGTGAGT GGCGGATGGG	780
	CGGTTGTCCG TCTGTAACAA TGGATGCGTG TGCATTATTA CAAAAATTCG ACTTTTGTAA	840
	TAATATTTCA CATTTCGAC ACTTTTTTGC TATAAACAA CCAATTGAGC GATAATAAAT	900
15	TCGCTTTTAA AAAATATGAG TTATCTATTT AGTTGCCAA GATAAAATAA TAATGTTTAA	960
	TAACATCATA TAGAGTATGT TAGTTTTTAA TGTGGAATAT ACGAATGTGc AAACAAAGTA	1020
	ATCGGTAGAA ATTCAACATA CATAGCGCCG TTTACTGTTA AGTATTCACA TTACAGATGA	1080
20	AAAATATAAA ATTCTACATA ATCAAGACCA TGATGTGTAC TTGTTTAACT TATGACTCTA	1140
	TTTGTTTAAC AATTGCGATA ATGGTCTTTT TATTTTATGC GTATCATTCTG TCATATTTTT	1200
25	TATGAGGAAG GAGAAATGAT TATGTTAAGT ATTAAGCATT TAACGAAAAT TTATTCTGGT	1260
	AATAAAAAGG CAGTAGATGA CATCTCTTTA GATATTCAAT CTGGGGAATT TATCGCATT	1320
	ATTGGAACCA GTGGAAGTGG CAAAACGACT GCTTTAAGAA TGATAAACCG TATGATTGAA	1380
30	GCGACAGAAG GACAAATTGA AATTGATGGT AAAGATGTTT GGAGTATGAA TCCTGTGCGAA	1440
	TTGCGTAGAA ATATTGGCTA TGTTATTCAA CAAATTGGCT TAATGCCTCA TATGACGATT	1500
	AAAGAGAATA TTGTGTTGGT ACCCAAATTG TTGAAATGGA CTAAAGAGGA AAAGGATAAA	1560
35	CGTGCAAAGG AATTAATTAA ACTTGTGGAT TTACCGGAGT CATTTTTAGA GCGTTATCCA	1620
	GCAGAACTAT CAGGTGGGCA ACAACAACGT ATCGGTGTTG TAAGAGCACT TGCGGCCGAA	1680
	CAAGATATTA TTTAATGGA TGAACCTTTT GGTGCATTGG ATCCTATTAC GAGAGATACG	1740
40	TTACAAGATT TAGTTAAAAC GTTACAACGA AAATTAGGCA AGACGTTTAT CTTTGTAAAC	1800
	CATGATATGG ATGAAGCGAT TAAATTAGCA GACAAAATTT GTATTATGTC AGAAGGTAAG	1860
	GTGGTGCAAT TTGATACGCC AGACAATATT TTAAGACATC CCGCAAATGA TTTTGTACGT	1920
45	GATTTTATAG GACAAAATAG ACTGATTCAA GACCGTCCCA ATGACAAGAC TGTAGAAGGT	1980
	GTAATGATTA AACCAATCAC GATACAAGCA GAAGCAACAC TGAATGACGC CGTTCATATT	2040
50	ATGAGACAAA AACGTGTTGA TACTATTTTT GTAGTAGATA GTAATAACCA TTTACTAGGT	2100
	TTCTTAGACA TTGAAGATAT AAATCAGGGT ATACGTGGAC ACAAAGTTT ACGAGACACC	2160

55

## EP 0 786 519 A2

	ATTTTAAAAA GAAACGTTAG GAATGTACCT GTCGTAGATG ATCAACAGCG TTTAGTAGGA	2280
	CTGATTACGC GTGCCAATGT TGTGATATT GTATATGACA CGATTGGGG CGATAGTGAG	2340
5	GATACAGTGC AAACAGAACA TGTGGGGGAA GACACTGCGT CCTCAAAAGT GCATGAGCAA	2400
	CACACTACTA ATGTCAAAGT ACGTGACATA GGAGATGATA AATCATGATT GAGTTCCTAC	2460
	ATGAACATGG TGGACAGTTG ATGTCGAAAA CACTGGAACA TTTCTATATT TCTATAGTGG	2520
10	CATTATTACT TGCCATCATT GTTGCACTAC CTATAGGCAT TTTATTATCA AAAACAAAGC	2580
	GAACGCGCAA TATTGTATTA ACTGTGGCAG GTGTCTTACA AACTATTCCA AACTAGCTG	2640
15	TACTTGCTAT TATGATACCG ATTTTGGTG TTGGTAAAC GCCTGCAATT GTACGCTAT	2700
	TTATTTATGT ATTATTACCT ATTTTAAATA ACACGGTACT CGGTGTTCAA AATATTGATA	2760
	GCAACATTAA AGAAGCTGGA AAAAGTATGG GAATGACACA ATTTCAATTG ATGAAGGATG	2820
20	TTGAATTGCC GTTAGCATTG CCGCTTATCA TTGGTGGCAT TCGTTTGTCA TCTGTGTATG	2880
	TAATTAGTTG GGCTACACTT GCAAGTTATG TAGGTGCGGG TGGATTAGGT GATTTCATTT	2940
	TCAATGGTTT AAATTTATAT GATCCACTGA TGATTGTAAC TGCAACGGTA CTCGTTACTG	3000
25	CACTAGCATT AGGTGTTGAT GCCTTATTAG CTTTAGTTGA AAAATGGGTA GTTCCCAAAG	3060
	GCTTAAAGT ATCTGGATAA TTAGGAGGCT AAGATAATGA AGAAAATTAA ATATATACTT	3120
	GTCGTGTTTG TCTTATCGCT TACCGTATTA TCTGGATGTA GTTTGCCCGG ACTAGGTAGT	3180
30	AAGAGCACGA AAAATGATGT CAAAATTACA GCATTATCAA CAAGCGAATC GCAAATTATT	3240
	TCACATATGT TACGGTTGTT AATAGAGCAT GATACACAG GTAAGATAAA GCCAACATTA	3300
	GTAAATAATT TAGGGTCAAG TACGATTCAA CATAATGCCT TAATTAATGG GGATGCTAAT	3360
35	ATATCAGGTG TTAGATATAA TGGCACAGAT TTAACGGGAG CTTTGAAGGA AGCACCAATT	3420
	AAAAATCCTA AGAAAGCAAT GATAGCAACA CAACAAGGAT TTAAAAAGAA ATTTGATCAA	3480
	ACGTTTTTTG ATTCGTATGG TTTTGCGAAT ACGTATGCAT TCATGGTAAC GAAGGAAACC	3540
40	GCTAAAAAAT ATCATTTAGA GACAGTTTCA GATTTAGCAA AGCATAGTAA AGATTTACGT	3600
	TTAGGTATGG ATAGTTCATG GATGAATCGT AAAGGCGATG GCTATGAAGG ATTTAAAAAA	3660
45	GAGTATGGTT TTGACTTTGG TACAGTGAGA CCAATGCAAA TAGGTCTAGT CTACGACGCA	3720
	TTAAACTCAG AGAAGTTAGA CGTTGCATTA GGTATTCTA CAGATGGTCG AATTGCGGCG	3780
	TATGATTGTA AAGTACTTAA AGATGATAAA CAATTTTCC CACCTTATGC TGCGAGTGCT	3840
50	GTTGCAACAA ATGAATTATT ACGGCAACAC CCAGAACTTA AAACGACGAT TAATAAGTTG	3900
	ACAGGAAAGA TTTCGACTTC AGAGATGCAA CGCTTGAATT ATGAAGCGGA TGGTAAAGGT	3960

55

AAAGGTGGTC ATAAGTAATG GAAGGTAATT TATTACAGCA ATTATTCAAT TATTATGTTA 4080  
 CGAACTTTGG TTATCTATGG GATTTATTTT TCAAACACTT ATTAATGTCT GTCTATGGTG 4140  
 5 TGCTGTTTGC AgCTTTAATT GGTATTCCAT TGGGAATCTT GCTTGCaAGA TACACAAAAC 4200  
 TTTCTGGATT TGTAATTACA ATTGCAAATA TAATTCAAAC AGTTCCAGTC ATTGCAATGT 4260  
 TAGCTATTTT AATGTTAGTC ATGGGCTTAG GTTCAGAAAC AGTAGTTTTA ACAGTGTTTT 4320  
 10 TATATGCGTT ACTTCCAATT ATAAAAACA CTTATACTGG TATAGCTAGT GTTGATGCGA 4380  
 ATATTAAGGA TGCTGGCAAA GGTATGGGAA TGACACGCAA TCAAGTGCTA CGAATGATTG 4440  
 AATTACCGTT ATCTGTTTCG GTTATTATCG GTGGCATTCTG TATTGCCTTG GTTGTTGCGA 4500  
 15 TAGGTGTTGT TGCCGTTGGA TCATTTATAG GAGCACCTAC GCTTGGTGAC ATTGTGATTC 4560  
 GTGGTACAAA TGGACGGAT GGCACAACGT TTATTTTAGC AGGTGCGATT CCGATTGCTA 4620  
 TCATTGCAAT CGTCATTGAT GTACTATTAA GATTTTATAG AAAACGATTA GACCCAACAA 4680  
 20 CACGACATCG TAAAAATCAA TCTAATCATC GGCCGCAAG TATTAATATG TAATAGTAGA 4740  
 AGATGTTTAT AATTTAGCGA TTTCGTTTCA TGATTATATA AAAATGAGGC TACTCAAGGA 4800  
 25 GCTCAAATAA TCTTTGAGTA GCCTTTTTAT AGGTGTGTGT TGTATGCGTT TACACTAAAA 4860  
 TAGCAATTAT TATCATGAAA GTTTTTGGAT AAAAAGCGTT AATTATTGTA AAAATACTAA 4920  
 AAAATGAGAT GTTTTATTTA TAATTTTCTG CAAATTTATG ATATTGTTTC TTAATATATC 4980  
 30 ATATTAAAAA TTTGTTTTTC TTAAACATAG GAGGCTTATC TAATTCATGG ACACATCAAA 5040  
 ACAATTTAGA GGTGACAACC GATTGCTTTT GGGTATCGTT TTAGGGGTTA TTACCTTTTG 5100  
 GCTATTCGCG CAGTCACTTG TTAATCTTGT TGTCCCATTA CAATCAACAT ATAGTAGTGA 5160  
 35 CGTTGGAACG ATAAATATCG CTGTTAGCTT ATCTGCCTTA TTTGCTGGTT TGTTTATCGT 5220  
 AGGTGCTGGT GATGTTGCTG ATAAATTGG TCGCGTCAA ATTACTTATG TAGGATTGAT 5280  
 ATTAATGTT GTAGGTTTCAT TACTCATCAT CATTACACCT TGCCAGCAT TTTAATTAT 5340  
 40 AGGTAGAATA ATTCAAGGTT TGTCTGCAGC ATGTATTATG CCATCAACAC TTGCTATTAT 5400  
 TAACGAATAT TATATTGGTA CAAGAAGACA ACGTGCTTA AGCTATTGGT CTATTGGTTC 5460  
 TTGGGGTGGT AGTGGTATTT GTACGTTGTT TGGTGGCTTA ATGGCTACAT ATATAGGTTG 5520  
 45 GCGTTCAATA TTTGTTGTTT CAATTCTATT AACATTATTA GCAATGTACT TAATCAAACA 5580  
 TGCACCTGAG ACTAAAGCAG AACCAATCAA AGGTATGAAA GCAGAAGCTA AAAAGTTTGA 5640  
 50 CGTTATTGGT TTAGTCATTT TAGTAGTGAC GATGTTAAGT TTAAATGTAA TCATCACACA 5700  
 GACGTCTCAT TTTGGTTTAG TTTCACCGTT AATTCTAGGT TTAATTGTTG TGTTTATCTG 5760

55

	AATTTTAAA AATAGAGGAT ACAGTGGTGC AACTATTTCA AACTTCTTAT TAAATGGTGT	5880
	AGCAGGTGGT GCACTTATCG TTATTAACAC GTATTATCAA CAACAATTAG GATTTAATTC	5940
5	TTCCGAAACG GGTATATTT CATTAACGTA TTTAATAACA GTGTTGTCAA TGATTCGTGT	6000
	AGGTGAAAAG ATTTTATCTC AACATGGTCC GAAGCGCCCA CTATTACTAG GAAGTGGCTT	6060
	TACAGTGATT GGGTTAATCT TATTGTCGTT AACATTTTCA CCAGAAGTGT GGTATATCAT	6120
10	ATCTAGTATA GTTGGATATT TATTGTTTGG TACTGGTTTA GGATTATATG CTACACCATC	6180
	AACTGATACA GCAGTTGCTA GTGCGCCAGA TGATAAGTCG GGTGTTGCTT CAGGTGTGTA	6240
	TAAAATGGCG TCATCATTAG GAAATGCATT TGGAGTAGCA GTATCTGGTA CGGTTTATAC	6300
15	TGTGTTAGCA GCTAATTTAA ATTTGAACIT AGGTGGTTTC ACAGGTATGA TGTTTAATGC	6360
	CTTGCTAGCA ATTGTTGCAT TTTTAGTCAT TTTACTATTA GTTCCTAAAA ATCAAACGAA	6420
20	TTTGTAAGAC TGAAATGAAA GCAAGTTATT ATGTAGGGAT TTTAAAGGAA ATTTTGTGAA	6480
	AGTAAGTTTA TCATACACAC TTAATGTTGC GTATTGACGT TTAATGTTAG GTGTGTTCTT	6540
	TTATAGACGA TAAAAGCTGT GTGCATATTA AGCGAATGAT TTTCAAATTG ACGCTAATAT	6600
25	GCGAAAGTAG TATTTTAAA ATGAACAACA ACGATGAAGA GGGGTTTATA GGATGAAAAT	6660
	TGCAATTGCT GGATCGGGTG CATTAGGTAG TGGCTTTGGT GCCAACTAT TTCAAGCAGG	6720
	ATATGATGTC AACTTTATTG ACGGATATAC ATCTCATGTT GAAGCGGTTA AGCAACATGG	6780
30	ATTAAATATA ACGATTAATG GAGAGGCATT CGAGTTAAAC ATTCCGATGT ATCATTTTAA	6840
	TGATCAACCG GACGAAAGCA TTTACGATGT TGTCTTTCTA TTTCCAAAGT CTATGCAATT	6900
	AAAAGAAGTG ATGGAAGATA TGAAGCCACA TATTGATAAT GAAACGATCG TCGTATGTAC	6960
35	GATGAATGGT CTGAAGCATG AAGAAGTCAT TGCGCAGTAT GTTGCTCAAT CACAAATTGT	7020
	CAGAGGTGTT ACGACTTGGG CGGCAGGTCT TGAAAGCCCT GGACACAGTC ATTTACTTGG	7080
	TAGTGGACCA GTTGAAATAG GTGAACTAGT GGATGAAGGT AAAGAAAATG TTATAAAAGT	7140
40	TGCTGATTTA CTTAACGAAG CGGAATTGAA TGGTGTCAAT AGTAAAGATT TATACCAATC	7200
	GATTTGGAAG AAGATTTGTG TTAATGGTAC GGCAAATGCA TTAAGCACAG TGTGAGAGTG	7260
	TAATATGGCA TCGCTGAATG AAAGTAGTTA TGCGAAGTGT TTGATTTATA AATTAACGCA	7320
45	AGAAATAGTG CATGTAGCGA CGATTGATAA TGTTCAATTA AATGTTGATG AAGTATTTGA	7380
	ATATTTAGTT GATTTAAATG AAaAAGTTGG TGCGCATTAT CCATCCATGT ATCAAGATTT	7440
50	AATTGTTAAT AATAGAAAAA CTGAAATTGA TTATATTAAT GGCGCAGTTG CAACATTAGG	7500
	TAAACAACGT CaTATTGAAG CGCCAGTCAA TCGCTTTATT ACTGATTTAA TTCATACTAA	7560

55

	CAATCACGTG ATATTACGGT CATTATTAAG ATTGAAATGT AATAAATAAA GAACAGCAGT	7680
	AAGGTACTTT CAAATTGAAA TGATCTTGGT GCTGTTTTTC TTGATTGATC TTCGTCATAA	7740
5	TTCAGATTTC TCATAGGcTA CGACATACTA TTAGTATTTA CTAGACAGTT TTTACGACGA	7800
	CACTTTGAAA AATTTTGAGG CAAATCATTT GGAAGTCTCA CGTGAATTTT GTAAACTCAT	7860
	CAAGCAAGTA ATTATATTAA AAAGACAAAT AGAGAAAAGG TGTTTATAAT GAGTAAAATT	7920
10	TTTGTAACGT GTGCAACGGG CCTTATTGGC ATTAAATTAG TTCAAAGACT AAAAGAAGAG	7980
	GGGCATGAGG TTGCTGGTTT TACTACATCT GAGAATGGTC AACAAAAGCT AGCTGCTGTT	8040
	AATGTAAAAG CATATATTGG TGATATATTA AAAGCTGATA CTATTGATCA AGCGTTAGCA	8100
15	GATTTTAAAC CAGAAATCAT TATCAATCAA ATTACGGATT TAAAAAATGT TGATATGGCA	8160
	GCAAATACGA AAGTACGTAT TGAAGGTTCT AAAAACCTAA TTGATGCGGC GAAAAAGCAT	8220
	GACGTTAAGA AAGTAATTGC CCAAAGTATT GCCTTTATGT ATGAACCTGG CGAAGGATTA	8280
20	GCAAATGAGG AACTTCACT TGATTTTAAAC TCAACTGGCG ATAGAAAAGT AACGGTTGAT	8340
	GGTGTGGTTG GTTTAGAAGA AGAAACGGCT CGTATGGATG AATACGTTGT TTTACGTTTT	8400
25	GGCTGGTTAT ATGGCCCAGG TACTTGGTAC GGAAAAGATG GCATGATTTA TAATCAATTT	8460
	ATGGATGGTC AAGTGACACT TTCAGATGGC GTAACATCAT TTGTGCATCT TGATGATGCA	8520
	GTTGAAACAT CTATTCAAGC TATTCATTTT GAAAATGGTA TCTATAATGT AGCAGATGAT	8580
30	GCACCTGTTA AAGGTTCTGA ATTTGCAGAA TGGTATAAAG AACAACTTGG TGTTGAACCA	8640
	AATATTGATA TTCAACCTGC GCAACCATTT GAACGTGGCG TAAGCAATGA GAAGTTTAAA	8700
	GCGCAAGGTG GTACTCTGAT TTATCAAAC TGGAAAGATG GCATGAATCC AATTAAATAA	8760
35	TAATTTATCC GTTTAATATA CAAAGAATAA AGACTTGGTC GAATCGTGA TGATATATTA	8820
	TCAAACGCAC GGCTCGAACA AGTCTTTTTT ATTATGTCTT CGTTATCTTT GTATGAAGGA	8880
	ATAACAGAAT TACAATTAAT GTACTGAATA ATGCAATTAA TGTTGTGATT AGTGCTAATT	8940
40	TAATTTCTAT TGGTAGCCAA GTCAGTACAA AAGACCAATT ATTGCTACCG AGAATGAGAT	9000
	ATGGTAATGC ATATAATATG AGCGCTAAAG CGATACATAT ACATAATGAT AACCAACTCA	9060
45	ATACAGCAAT CC	9072

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16826 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

EP 0 786 519 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

	GTGGAACAGC TGTAACATA TCATTCTTTT CAACATTTAT TGGGAAAATG TTAGCTACAT	60
5	TTCTATATCC GATTAATAAT GTAGACTTTT CATATATnTC TGTAATGAA AGTGACAATA	120
	TAAAGAAGCA ATATTTGAAA ACTAATCTAA TTGCTATAGC TGCCCTATGT TTAGTCATGA	180
	TTATATGTTA TCCAATTACA ATAATTATTG TCTCTTTACT GTATAACATT GATTCAAGTT	240
10	TATATTCGAA GTTTATTATT TTAGGTAATA TAGGTGTTTT ATTCAATGCA GTGAGTATTA	300
	TGATCCAAAC TTAAATACA AAACACGCAT CAATAACATT ACAAGCGAAT TATATGACGC	360
	TTACACGAT TACATTTATA TTCATAACTA TTTTAATGAC AATTGCGTTT GGTCTAAATG	420
15	GATTCTTTTG GACAACGCTG TTCAGCAACA TTATTAAGTA TGTGATTTTA AATATTATAG	480
	GTTTAAAGTC TAAATTCATT AATAAAAAGG ACGTCGATTA GATGAGTGAA AAAAAGATTT	540
	TGATTTTATG TCAGTATTTT TATCCGGAAT ATGTATCTTC TCGACGTTA CCAACTCAAT	600
20	TGGCGGAAGA TTAAATTGCG AATCACATTA ATGTCGATGT CATGTGTGGA TGGCCATATG	660
	AATATAGTAA TCATAAACAG GTTCTAAAA CCGAGATGCA TCGTGGTATT CGCATTCGAC	720
25	GTCTCAAGTA TTCGAGGTTT AATAACAAAA GTAAGGTTGG AAGGATCATC AATTCTTTTA	780
	GTTTATTTTC AAAATTCGTG ATTAATATAC CTAAAATGTT GAAATATGAT CAGATTCTTG	840
	TTTACTCTAA TCCACCAATC TTGCCATTAA TACCAGACGT TTTACACAGA CTGCTTAAGA	900
30	AAAAATATTC TTTTGTGGTG TATGATATAG CACCTGATAA TGCATTAAAG ACAGGTGCAA	960
	CTCGTCCAGG TAGCATGATT GATAAGCTGA TGCCTTACAT TAATAGACAT GTCTACAAGA	1020
	ATGCTGAAAA TGTCAATGTC CTTGGTACGG AAATGAAAAA CTACTTACTA AATCATCAAA	1080
35	TTTCTAAAAA TGCTGACAAT ATCCATGTGA TTCCTAACTG GTATGACATG CGTCAATTAC	1140
	AAGACAATCG TATCTATAAT GACACATTTA AAGCTTACCG TGAGCAATAC GACAAAATTT	1200
	TATGTATAG CGGTAATATG GGGCAGTTAC AGGATATGGA GACACTTATC TCATTTTTAA	1260
40	AATTAAATAA GGATCAGTCT CAAACGTTAA CAATACTTTG TGGTCATGGT AAGAAATTTG	1320
	CAGATGTCAA AACGGCAATA GaAGACCATC GTATTGAAAA TGTTAAAATG TTTGAGTTTT	1380
	TAACAGGTAC AGACTATGCT GACGTATTAA AAATTGCGGA TGTATGTATT GCATCGCTGA	1440
45	TTAAAGAAGG CGTCGGTTTA GCGTGCCGA GCAAGAAITA TGGCTATCTT GCAGCTAAGA	1500
	AAGCGTTGGT ACTCATCATG GATAAGCAAT CTGATATCGT TCAACATGTT GAACAATATG	1560
50	ATGCGGGTAT CCAAATTGAT AATGGCGATG CACATGCCAT TTATAACTTC ATCAACACTC	1620
	ACTCGAGTAA GGAATTGCAC GAGATGGGTG AGCGCGCACA TCAACTGTTT AAAGATAAAT	1680

55

EP 0 786 519 A2

	AAGCGATTAT TCGATGTAGT GAGTTCAATA TATGGTTTAG TAGTTTAAAG TCCGATTCTG	1800
	TTAATTACAG CATTACTAAT TAAAATGGAA TCACCTGGAC CAGCCATTTT CAAACAAAAA	1860
5	AGACCGACGA TTAATAATGA ATTGTTTAAT ATTTATAAGT TTAGATCAAT GAAAATAGAC	1920
	ACACCTAATG TTGCAACTGA TTTAATGGAT TCAACATCGT ATATAACAAA GACAGGGAAG	1980
	GTCATTCGTA AGACCTCTAT TGATGAATTG CCACAATTAT TGAATGTTTT AAAAGGAGAA	2040
10	ATGTCAATTG TAGGTCCTAG ACCAGCGCTT TATAATCAAT ACGAATTAAT CGAAAAACGT	2100
	ACAAAAGCGA ACGTGCATAC GATTAGACCA GGTGTGACAG GACTAGCTCA AGTGATGGGG	2160
	AGAGATGATA TCACTGATGA TCAAAAAGTA GCGTATGATC ATTATTACTT AACACATCAA	2220
15	TCTATGATGC TTGATATGTA TATCATATAT AAAACAATTA AAAATATCGT TACTTCAGAA	2280
	GGTGTGCATC ACTAATGAGA AAAAATATTT TAATTACAGG CGTACATGGA TATATCGGTA	2340
	ATGCTTTAAA AGATAAGCTT ATTGAACAAG GACATCAAGT AGATCAAATT AATGTTAGGA	2400
20	ATCAATTATG GAAGTCGACC TCGTTCAAAG ATTATGATGT TTTAATTCAT ACAGCAGCTT	2460
	TGGTTCACAA CAATTCACCT CAAGCAAGGC TATCTGATTA TATGCAAGTG AATATGTTGC	2520
25	TGACGAAACA ATTGGCACAA AAGGCTAAAG CTGAAGACGT TAAACAATTT ATTTTATGA	2580
	GTACTATGGC AGTTTATGGA AAAGAAGGTC ATGTTGGTAA ATCAGATCAA GTTGATACAC	2640
	AAACACCAAT GAACCTACG ACCAACTATG GTATTTCCAA AAAGTTCGCT GAACAAGCAT	2700
30	TACAAGAATT GATTAGTGAT TCGTTTAAAG TAGCAATTGT GAGACCACCA ATGATTATG	2760
	GTGCACATTG CCCAGGAAAT TTCCAACGGT TAATGCAATT GTCAAAGCGA TTGCCAATCA	2820
	TTCCAATAT TAACAATCAG CGCAGTGCAT TATATATTAA ACATCTGACA GCATTTATTG	2880
35	ATCAATTAAT ATCATTAGAA GTGACAGGTG TGTACCATCC TCAAGATAGT TTTTACTTTG	2940
	ATACATCGTC AGTAATGTAT GAAATACGTC GCCAATCACA TCGTAAAACG GTATTGATCA	3000
	ACATGCCTTC AATGCTAAAT AAGTATTTTA ATAAGTTGTC GGTCTTTAGA AAATTATTCG	3060
40	GCAATTTAAT ATACAGCAAT ACGTTATATG AAAATAATAA TGCATTGAA ATTATTCCTG	3120
	GAAAAATGTC ACTTGTTATT GCGGACATCA TGGATGAAAC GACAACCAA GATAAGGCAT	3180
	AAGTCATCTA TTAAATAAAA TCAACATACA AATCGTTTTA TTTGGAGGTT ATAGTATGAA	3240
45	GTTAACAGTA GTTGGCTTAG GTTATATTGG TTTACCAACA TCAATTATGT TTGCAAAACA	3300
	TGGcGTCGAT GTGCTTGGTG TTGATATTAA TCAGCAAACG ATTGATAAGT TACAAAGTGG	3360
50	TCAAATTAGT ATTGAAGAAC CTGGATTACA AGAGGTTTAT GAAGAGGTAC TGTCAATCGG	3420
	AAAATTGAAG GTATCTACAA CGCCAGATGC ATCTGATGTT TTTATCATTG CCGTTCCGAC	3480

55

	TAGTATTTTA TCATTTTITAG AAAAAGGAAA TACCATTATT GTAGAGTCGA CAATTGCGCC	3600
	TAAAACGATG GATGATTTTG TAAAACCACT CATTGAAAAT TTAGGGTTTA CAATAGGTGA	3660
5	AGATATTTAT TTAGTGCACT GTCCAGAACG TGTACTGCCA GGAAAAATTT TAGAAGAATT	3720
	AGTTCATAAC AATCGTATCA TTGGCGGTGT GACTGAAGCT TGTATTGAAG CGGGTAAACG	3780
	TGTCTATCGC ACATTCTGTC AGGGAGAAAT GATTGAAACA GATGCACGTA CTGCTGAAAT	3840
10	GAGTAAGCTA ATGGAAAACA CATATAGAGA CGTGAACATT GCTTTAGCTA ATGAATTAAC	3900
	AAAAATTTGC AATAACTTAA ATATTAATGT ATTAGATGTG ATTGAAATGG CAAACAAACA	3960
	TCCGCGTGTT AACATCCATC AGCCTGGTCC AGGTGTAGGC GGTCAATGTT TAGCTGTTGA	4020
15	TCCGTACTTT ATTATTGCTA AAGACCCTGA AAATGCAAAG TTAATTCAAA CTGGACGTGA	4080
	AATTAATAAT TCAATGCCGG CCTATGTTGT TGATACAACG AAGCAAATCA TCAAAGTGTT	4140
	GAGCGGGAAT AAAGTCACAG TATTTGGTTT AACTTATAAA GGTGATGTTG ATGATATAAG	4200
20	AGAATCACCA GCATTTGATA TTTATGAGCT ATTAAATCAA GAACCAGACA TAGAAGTATG	4260
	TGCTTATGAT CCACATGTTG AATTAGATTT TGTGGAACAT GATATGTCAC ATGCTGTCAA	4320
25	AGACGCATCG CTAGTATTGA TTTTAAGTGA CCACTCAGAA TTTAAAAATT TATCGGACAG	4380
	TCATTTTGAT AAAATGAAGC ATAAAGTGAT TTTTGATACA AAAAATGTTG TGAAATCATC	4440
	ATTTGAAGAT GTATCGTATT ATAATTATGG CAATATATTT AATTTTATCG ACAAATAAAA	4500
30	TGTGTCAAAC TAGGGCATAC ATGATTAAGG AAAGATAAGC TGTCATGTGT TTGAACTTCA	4560
	GAGAGGATAA TGTTATGAAA AAAATTATGG TTATTTTCGG TACGAGACCC GAAGCAATAA	4620
	AAATGGCACC ATTAGTAAAA GAAATGATC ATAATGGGAA CTTTGAAGCG AACATTGTGA	4680
35	TTACAGCACA ACATAGAGAT ATGTTAGATA GTGTGTTAAG TATATTTGAT ATTCAAGCTG	4740
	ATCATGATTT AAATATTATG CAAGATCAAC AAACATTAGC AGGCCTTACG GCGAATGCAC	4800
	TTGCTAAACT TGATAGCATC ATTAATGAGG AACAAACCGA TATGATTTTA GTACATGGTG	4860
40	ATACTACAAC GACTTTTGTA GGAAGTTTGG CAGCATTTTA TCATCAAATT CCGGTCGGAC	4920
	ATGTAGAAGC TGGACTTCGA ACACATCAGA AATACTCACC ATTTCTGAA GAGTTAAATC	4980
45	GAGTCATGGT AAGTAATATT GCTGAATTGA ATTTTGCGCC AACAGTAATT GCAGCTAAAA	5040
	ATTTACTTTT TGAAAACAAA GACAAAGAGC GTATCTTTAT TACTGGAAAT ACAGTTATTG	5100
	ACGCATTGTC AACAAACAGT CAAAATGATT TTGTTTCAAC GATTATTAAT AACATAAAG	5160
50	GCAAGAAAGT TGTTTTACTA ACAGCGCATC GTCGTGAAAA TATTGGGGAA CCGATGCATC	5220
	AGATTTTTTAA AGCAGTAAGA GATTTGGCAG ATGAATATAA AGATGTTGTC TTCATTTATC	5280

55



	GGATTGAATT AATTGAGCCA TTAGATGCGA TTGAGTTCCA TAATTTTACA AATCAATCGT	5400
	ACCTCGTGCT GACAGATTCT GGTGGTATTC AAGAGGAGGC TCCTACATTT GGAAAACCTG	5460
5	TGTTGGTATT AAGGAATCAT ACAGAGCGTC CCGAAGGCGT TGAGGCGGGA ACATCGAGAG	5520
	TAATTGGCAC AGATTATGAC AATATTGTTT GAAATGTGAA ACAATTGATT GAGGATGATG	5580
	AAGCGTATCA ACGTATGAGT CAAGCGAATA ATCCATATGG TGATGGACAA GCATCACGAC	5640
10	GTATTTGTGA AGCAATAGAA TATTATTTTG GATTGCGCAC AGACAAGCCG GATGAATTCG	5700
	TACCTTTACG TCACAAATAA TAAAAAACCC CTAATCATGA AGTTGGTTTA GACAACCAGC	5760
15	GGTGACTAGG GGTTTTTAAT ATATTTATTT TTGATAGTGG TAGCCAATAT CATATTTGAA	5820
	TACTTTATTT GATAATATTG GACTTTGCTG TCCATCGTCA TCACTTTTTA AACGTACATT	5880
	TTTATGAGCT TCTTTAAATA CATCGGAATT CAACCAATTA TTAAAGCTAT CTTCAGATTC	5940
20	CCAAATAGTT AAGATTTTAA CTTCTGCTGT ATCCTCGGTA TTTAATGTTT TAGTGACAAA	6000
	CATTTGTTGG AAGCCTTCAA TAGTTTCAAT ACCTTGCTCA TTGTAAAAAC GTTCAATCGT	6060
	TTCTTCCGCA CTGCCTTTTT GTAATTGTAA TCTATTTTCT GCCATAAACA TGGGCAATCA	6120
25	CTCCTCTATT TTATGATTG ATTGGGGTAA TGTTTTTACA AATGTAAAGA GTACAGCGGT	6180
	TTGTATGATA ACCATTATGA TTAATCCTAC ACGGACTGCA AGAACATCCA CCATATAAAT	6240
	TGAAAAACCT ATTACAATGT ATAAGCTAAT TAAAATTTTA ATTTTCTGTT GTAGCGTGTA	6300
30	GCCTCGATGT AAATAAAAGT TTTCTACATA TTCTTTATAA ATTTTTTGAT TAATAAGCCA	6360
	ATTGTAAAAG CGATCTGAAC TTCGAGCAAA GCAAAAAACT GCTACGAGTA AAAAAGGGGT	6420
	CGTTGGCAGT AAAGGTAATA CGGCACCTGC AATACCAAGC GCTGTAAATA TTAAGCCAAT	6480
35	GACGATTAAA ATAAGTCGCA TTGAAAAAAC TCCATTCTAG TACTAATGCG CATGTAATAT	6540
	TGTTTTAGTA ATATAACTCA TGCTAAATAT AATGTGTATG ATAAGTGCAA TGACTCAGTA	6600
	AAATGAAACG ATGTTGAATT ATCCTTGTC AATTAACGCA TTTTAAGCGC GACTTTTCATA	6660
40	ACAACCAAAC TATTTAATGA GAATTATTCT CAAGTATTAT AGTTATATTA TGTGTTTTAT	6720
	TTTGAAAAG TGCAATATGT TTTGAAAAAT AAGATTATTT TTATGTGCAA AAACGACGCA	6780
45	AAAGTTTTAA AAATGAGACT TCTGTGAGCT GATTATTTTA TAAAATGTAA ACGCTTACTA	6840
	TATAATGTGA ATCATATCGT TTAAAGCAT TATTAAATAT GATGCTAAGA GATTTATATT	6900
	ATAGCCAATA AACAAAGGAG AGATAATATG GCAGTAAACG TTCGAGATTA TATTGCAGAG	6960
50	AATTATGGTT TATTTATCAA TGGGGAATTT GTTAAAGGTA GCAGTGACGA AACAAATCGAA	7020
	GTGACTAATC CAGCAACTGG AGAAACACTA TCACATATTA CAAGAGCAAA AGATAAAGAT	7080

55

	TCAGAACGTG	CACAAATGTT	GCGTGATATT	GGTGATAAAT	TAATGGCACA	AAAAGATAAA	7200
	ATTGCAATGA	TTGAAACATT	AAATAATGGT	AAACCGATTG	GTGAGACAAC	AGCAATTGAT	7260
5	ATTCCATTTG	CTGCAAGACA	TTTCCATTAT	TTGCGAAGTG	TTATTGAAAC	AGAAGAAGGT	7320
	ACAGTGAATG	ATATCGATAA	AGACACAATG	AGTATCGTAC	GACATGAGCC	GATTGGCGTC	7380
	GTAGGTGCTG	TTGTTGCTTG	GAACTTCCCA	ATGCTATTAG	CTGCATGGAA	GATTGCGCCA	7440
10	gCCATTGCTG	CAGGTAATAC	AATTGTGATT	CAACCTTCGT	CTTCAACACC	ATTAAGTTTA	7500
	TTGGAAGTTG	CTAAAATTTT	CCAAGAGGTA	TTACCTAAAG	GTGTTGTCAA	TATACTAACG	7560
	GGTAAAGGTT	CAGAATCAGG	TAATGCAATT	TTCAATCATG	ATGGTGTAGA	TAAATTATCA	7620
15	TTTACGGGCT	CAACTGATGT	AGGTTATCAA	GTTGCCGAAG	CTGCAGCAA	ACATCTAGTA	7680
	CCCGCTACAT	TAGAGCTTGG	TGGTAAAAGC	GCCAATATCA	TATTAGATGA	TGCTAATTTA	7740
20	GACCTTGCAg	TTGAAGGTAT	TCAGTTAGGT	ATTTTATTCA	ACCAAGGTGA	AGTATGTAGT	7800
	GCAGGTCTC	GATTATTAGT	TCATGAAAAA	ATTTATGATC	AATTGGTGCC	ACGTTTACAA	7860
	GAGGCATTTT	CAAATATTAA	AGTTGGAAAT	CCACAAGATG	AAGCTACACA	AATGGGTAGT	7920
25	CAAAGTGGTA	AGGATCAATT	AGATAAAATT	CAATCATATA	TTGATGCAGC	AAAAGAATCA	7980
	GATGCACAAA	TTTTAGCAGG	CGGTCATCGC	TTAACTGAAA	ATGGATTAGA	TAAAGGGTTC	8040
	TTCTTTGAGC	CGACATTAAT	TGctGTGCCA	GACAATCATC	ACAAATTAGC	ACAAGAAGAA	8100
30	ATATTTGGAC	CAGTGTTAAC	AGTGATTAAA	GTGAAGGACG	ATCAAGAAGC	AATTGATATA	8160
	GCTAATGATT	CTGAGTATGG	TTTAGCAGGC	GGTGTATTTT	CTCAAAATAT	CACACGTGCA	8220
	TTAAATATTG	CTAAAGCTGT	ACGTACAGGA	CGTATTTGGA	TTAACACTTA	CAACCAAGTA	8280
35	CCAGAAGGCG	CACCATTGG	TGGTTATAAA	AAATCAGGTA	TCGGTCGAGA	AACTTATAAA	8340
	GGTGcGTTAA	GTAACATCA	ACAAGTTAAA	AATATTTATA	TTGATACAAG	CAATGCTTTA	8400
	AAAGGTTTGT	ACTAGAATAA	ATATCGTTTC	TGAAGCGTGT	TTGTAGGTCA	GTCTAGCGGT	8460
40	AAGTCTTAAC	ATTTAACGGC	GTTGTTTAGA	TTTTAAGCAA	AACAAAATAT	ATAGGAACAC	8520
	GTATCATGAT	ATTAGGATAT	AATGACTAAA	ATAATAGCAG	TAGGATGGTT	TTTAATTGCA	8580
45	AATCATCTTA	CTGCTGTTTT	TAATTATGCT	AATTTGCGAT	GCGGCTATTA	TAAGGACAGA	8640
	GTTGTTTATT	AATTATGGTG	ATTTAGAAAT	ATGAAGTTCA	ATATGCAAAG	TCATCGTTTG	8700
	TTTTAATATG	CGGAACAATC	ATTAAAGTTA	TTGCGATTTT	TTGAACTTAA	TGAAACTAAA	8760
50	CAATAAATTT	GAGATACTTT	TTTGTCATTT	TTATGTAACT	AACACAATAA	TCTCGTACAT	8820
	TATTAATAAT	TTCTATATGA	TAGGAATAAA	GCAAAGCGCG	AGTGTGCTGT	AAAAGTTTTTC	8880

55

	GATGATGTAT AAATCATGGT TAATTACGGA AGCATTAAATA TTAACCTGAG AAGCTATAAA	9000
	GAATTATTTT TAAAAGCGAC AATATTAAAT ACGACGCATT TATTTAGGAG TGGCAAACGT	9060
5	ATGAATGGGA AAAAGGCGAA TACGATAAAC AGATACAAAT ATTTTCATCA TGTCAATCAT	9120
	CAAAAAATTC AACAAAGTTC TAAAAAGACG CTGTGGGCAT CACTAATCAT CACATTGTTA	9180
	TTTACAGTGA TTGAATTTGT CGGAGGTTTA GTATCTAATC CATTGGCATT ACTGTCAGAT	9240
10	TCATTTTCATA TGCTTAGTGA TGTATTAGCA CTTGGTTTAT CTATGTTGGC CATTTATTTT	9300
	GCAAGTAAAA AGCCGACTGC ACGATACACA TTTGGATATT TAAGATTGGA GATATTAGCT	9360
	GCATTTTAA ATGGTTTAGC ATTAATTGTA ATTTCAATCT GGATTTTATA TGAAGCTATT	9420
15	GTACGTATTA TTTATCCGCA ACCAATTGAA AGTGGCATT TGTATTATGAT TGCTAGTATT	9480
	GGTTTACTCG TCAATATTAT TTTGACTGTT ATCCTTGTA GGTCTTTAAA ACAAGAAGAC	9540
	AATATCAATA TTCAAAGTGC ATTATGGCAT TTCATGGGAG ACTTATTGAA CTCTATTGGT	9600
20	GTCATCGTTG CAGTTGTATT GATTTACTTT ACAGGATGGC GCATCATCGA CCCAATCATT	9660
	AGTATTGTAA TTCACTCAT CATTTTACGT GGTGGTTATA AAATTACGCG TAATGCgTGG	9720
25	CTAATTTTAA TGGAAAGTGT GCCTCAACAT TTGGATACTG ATCAAATTAT GGCAGATATT	9780
	AAAAACATAG ATGGCATATT AGATGTACAT GAATTTTCATT TGTGGAGTAT TACAACAGAG	9840
	CATTATTCAT TAAGTGCCCA TGTTGTGTTA GATAAAAAAT ATGAGGGTGA TGATTATCAA	9900
30	GCGATTGATC AAGTATCATC ATTGTTGAAA GAAAAATATG GCATTGCACA TTCAACGTTG	9960
	CAAATTGAAA ACTTGCAATT GAATCCATTA GATGAGCCAT ACTTCGACAA ATTAACATAA	10020
	ATAAAACATT GTAGCGCCTA AAACATTAAAT CTATGTCATA GGCGCACGTT TCGTTTTATA	10080
35	CTTATGTTGC ATCATTTAAA TGATTTTCGT CAATTTCTTT GATGCTATCT ACATCTAACA	10140
	CGACATCTTT AGGTTTCAA ATATGAATAT GTTTTTCATC ATTTGTATGT AAAATGCGTT	10200
	CTATGATGTA CCTTTGACCG GCCATTGTTT CTACAGCAAT CTTTTGTGTT CTAGCTAAAC	10260
40	TTGCTACGAC AGATTCTTTA TCCATAATGA TAGCCCCCTA TATATATGTT TATTTACTTA	10320
	TACCCTAACA TGATTTTTAT ACTCTTTGAA AATATATTTT ACAGAATTTT ATCTAAATAT	10380
	TTAAAAAAT ATCTTAATAT CCTTGTAATC CGATAAGAAT TATAGTAATA TTTTTTCAAC	10440
45	CATcGTTATA GGAGGTCTTA TTAATGACAT TATTTTATT AGAAGCTAAC AATCTTGATT	10500
	TTGCATCAAC GAAAGAAGAA CTAGAAGCAA AGGCAGCATC ACTATCTACG AAGACAATTC	10560
50	CAACATTAAT TGAAGTACAA GCTACTGAAA ATTTAACTCA TGGTTATTTT ATTGTGGAAG	10620
	CAAATGACGA aGCAGAAGCT AAACAATTTT TAACAGAAGC AGATATTAGT ATTCAATTAG	10680

55

	TTGATTACCT TGTAACCTGG AACATTCCGG AAGGCATTAC GATGGATCAA TATTTAGCAC	10800
	GTAAAAAGAA AAATTCTGTT CATTATGAAG AAGTGCCAGA AGTTGAATTT AAACGCACAT	10860
5	ATGTATGTGA AGATATGTCT AAATGTATTT GTTTATACAA CGCACCTGAT GAAGAAGCGG	10920
	TACGTCGCGC GCGCAAAGCA GTTGATACAC CGATTGATGG CATCGAAAAA CTTTAATAAG	10980
10	ACAACAAGTT GATGAGATAT ATGTATATAG GTTTGGCATG GATTTGATT GCAGTTAATT	11040
	AGAATAGCTC AATGCTATAA ATGTAAGTAG TTGATATGAA GAACTAATG AACTAAATGC	11100
	AAGTATTGTC TAAAACAATC ATTTTATTGA AATTTAGTAG AGCTGAAATT AATATAACGT	11160
15	CGTTAATTGA ATAACGCTTA TGTATAAGA GCACTCATAC CAAACCATAA TCATCTATAG	11220
	ATATAACAAT TCACGATATA AGGGCTGTGT TTGGCATAGC CCTTTAGATA TACACTTAAT	11280
	TCCTATTAAA ATAGTAGGGA TTAAAGGGG GCTTGTCTAG ATTAAAATTC AACAAATTACA	11340
20	ACATCACTTT GGATCACATA AAGTAATTCA TAACTTTAAT TTGGACATTA GCAAGGGAGA	11400
	AATAGTCACT TTCATAGGGA AAAGTGGTTG CGGAAAGTCT ACTTTACTCA ATATTATCGG	11460
	TGGATTTATT CATCCATCGT CTGGTCGTGT CATTATTGAT AACGAAATTA AACAAACAGCC	11520
25	ATCTCCAGAT TGTTTAATGC TATTTCAACA TCATAATTG CTGCCATGGA AAACGATTAA	11580
	TGACAACATT AGGATTGGAT TACAACAGAA AATTAGTGAT GAAGAGATTA ACGCACAGCT	11640
	TAAATTAGTT GATTTAGAAG ACAGGGGAAA GCATTTTCCC GAGCAACTGT CCGGGGGTAT	11700
30	GAAACAACGT GTGGCACTAT GTCGAGCGCA TGTGCATAAG CCTAACGTTA TATTGATGGA	11760
	TGAGCCATTA GGTGCATTAG ATGCATTAC ACGTTATAAA CTTCAGGATC AACTAGTGCA	11820
	aCTAAAACAT AAAACGCAAT CAACTATTAT TTAGTGACG CATGACATTG ATGAAGCTAT	11880
35	TTATCTTTCC GACCGCATTG TTCTGTTAGG TGAAGGGTGC AATATTATTT CTCAATATGA	11940
	AATFACAGCA TCACATCCAC GCAGTCGTAA TGATAGCCAC CTACTTAAGA TTCGTAATGA	12000
40	AATTATGGAA ACATTTGCAT TGAATCATCA TCAAGTTGAA CCTGAATATT ATTTATAAGG	12060
	AGTGAGTGAC GATGAAAAGG TTAAGCATAA TCGTCATCAT TGGAATCTTT ATAATTACAG	12120
	GATGTGATTG GCAAAGGACG TCTAAGAAG GGTCTAAAAA TGCCCAAAT CAGCAAGTGA	12180
45	TTAAAATTGG ATATTTGCCG ATTACACATT CAGCTAATTT GATGATGACT AAAAAATTAT	12240
	TATCACAATA CAATCATCCG AAATATAAAC TAGAATTAGT TAAATTCAAT AATTGGCCAG	12300
	ATTTAATGGA CGCATTAAAC AGTGGTCGTA TTGATGGTGC ATCAACTTTA ATAGAGCTAG	12360
50	CGATGAAATC AAAACAGAAG GGCTCAAATA TAAAGGCTGT GGCATTGGGC CATCATGAAG	12420
	GCAATGTCAT TATGGGACAA AAAGGTATGC ACTTAAATGA ATTTAATAAT AATGGCGATG	12480

55

	GTAAACAATT AAAGATTAAA CCGGGGCATT TTAGCTATCA TGAAATGTCG CCAGCAGAAA	12600
	TGCCAGCCGC ATTGAGTGAA CACAGAATTA CAGGGTATTC TGTAGCCGAA CCATTCCGGT	12660
5	CACTGGGTGA AAAGTTAGGC AAAGGTAAGA CTTTGAAACA TGGTGATGAC GTTATACCTG	12720
	ATGCGTATTG CTGTGTGCTA GTACTGAGAG GGAATTGCT TGATCAACAC AAGGATGTAG	12780
	CGCAAgCATT TGTACAAGAT TATAAAAAGT CTGGCTTTAA AATGAATGAT CGCAAGCAAA	12840
10	GTGTAGACAT TATGACGCAT CATTTTAAAC AAAGTCGTGA CGTTTTAACA CAGTCAGCGG	12900
	CATGGACATC CTATGGTGAT TTAACAATTA AGCCATCCGG CTATCAAGAA ATTACGACAT	12960
15	TGGTAAAACA ACATCATTG TTTAATCCAC CTGCATATGA TGACTTTGTT GAACCGTCAT	13020
	TGTATAAGGA GGCATCGCGT TCATGACACG TCCACAAAT AACAAATTTA TATTACCTAT	13080
	TATCACATTT ATTATTTTCT TAGGCATTTG GGAAATGGTC ATTATTATTG GGCATTACCA	13140
20	ACCTGTATTG TTACCGGGTC CTGCTCTTGT AGGAAAAAGT ATATGGTCTT TCATTGTTAC	13200
	TGGAGAAATT TTCCAACATT TAGCAATTAG TTTATGGAGA TTTGTAGCGG GCTTTGTTGT	13260
	CGCATGTGTT GTTGCTATTC CATTGGGCTT CTTGCTTGGA AGGAATCGTT GGCTATACAA	13320
25	CGCTATCGAA CCGCTATTTC AATTGATTAG GCCGATATCT CCGATAGCAT GGCACCATT	13380
	TGTTGTTCTA TGGTTTGGTA TTGGTAGTTT GCCAGCGATT GCGATTATTT TTATCGCTGC	13440
	TTTTTTCCCA ATTGTGTTCA ATACTATTAA AGGCGTTAGA GACATTGAAC CTCAATATTT	13500
30	AAAAATAGCA GCAAATTTAA ATTTAACTGG GTGGTCATTG TATCGCAATA TATTATTTCC	13560
	CGGGGCATTT AAACAAATCA TGGCTGGGAT ACATATGGCG GTAGGAACAA GTTGGATATT	13620
	TTTAGTTTCT GGTGAAATGA TTGGTGACA ATCGGGATTA GGTTTTTTAA TCGTTGATGC	13680
35	ACGAAATATG TTGAACCTAG AAGATGTTTT AGCAGCAATA TTCTTTATCG GATTATTTGG	13740
	TTTTATTATT GATCGATTCA TTAGTTATAT TGAGCAGTTT ATACTTAGAA GATTGGTGA	13800
	ATAAGGAGAG ATGATGATGA CTTTAGAAAC GCTTATCAAA GAACAATTAG ATCCTCATTT	13860
40	AGTAGAAGTT GATGAAGGGA CGTATTATCC GAGAACATTT ATTCAGCAAT TATTGTAGA	13920
	TGGTTATTTT GGTGAGGCGG CATTGAGAAA AAATGCTGAA GTAATCGAAG CTGTATCGCA	13980
	GTCTTGTTTG ACAACAGGAT TTTGTTTATG GTGCCAATTA GCTTTTTCAA CGTATTTAGA	14040
45	AAATGCCACG CAGCCACATT TAAATAATGA CTTACAACAG CAATTGTTAT CTGGAGAAAT	14100
	ATTAGGTGCT ACCGGATTGT CTAATCCGAT GAAGTCATTT AATGATTTAG AAAAGTTGAA	14160
50	CCTTGAACAC ACTTATGTTG ATGGACAATT GGTGTGTCAGT GGACGTATGC CAGCTGTAAG	14220
	TAATATTCAA GAAGACCATT ATTTTGGTGC GATTTGAAAT CATGAATCAT CAGATGAATT	14280

55

	TTTAGGAGTC AACGGGTCAG CAACGTATCA AATCACATTG AATCAAGTCG TAGTGCCACA	14400
	ATCACAAATT ATCACGCATG ATGCGAAGCA GTTTGCGGCA ACTATTCGCC CGCAATTTAT	14460
5	TGCTTACCAA ATTCCAATAG GATTAGGCTC AATTAAAAGT TCTTTAGAGT TAATTGATGC	14520
	ATTTTCAAAT GTGCAAAACG GAATAAATCA ATATTTAGAG TATGATGTTG AAGCTTTTAA	14580
	AAAACGTTAT CGTCAACTTA GAGAGGAATA TTATGCAATA TTAGATGACG GTAACTTAAC	14640
10	TTCACATTTA AATGAATTAA TATCATTGAA GAAGGACATC GGCTATTTAT TGTTAGATGT	14700
	AAATCAAGCT TCTGTTGTCA ATGGTGGTTC TAGAGCGTAC ACACCATATT CGCCACAAGT	14760
	TCGCAAGTTA AAAGAAGGAT TCTTCTTCGC AGCATTGACA CCGACATTAA GACATTTAGG	14820
15	TAAACTTGAA GCAGAGTTGA AGGGGTAAGT GTGATAAGCT GATTTTTTGT TTAGATGCGT	14880
	TTGTTGAAAC ATTTTTTAAA ATAATATAAA TCTTAGTTTA TAAACATTTT CTGTTAATTT	14940
	GTTATATCCT TTAACTAGG AAAATATACA TTTCGTAATA ATAATAATCG TTATCATTGA	15000
20	AAAAGTGTTA ATAAGGTGTA TAATGAAAAT GTGAACAATT AATGAACTTC TTATTTTAAA	15060
	GAAGGTGAAT ACTATAGATA CGCATACTAA AGAACAACAA TTCTCGAATC TAGTAAGATC	15120
	TTATCGTAAA GAATACGTGG GTAAAGGACC CAATAGTATT CGAGTGTGCT TTAAGATAA	15180
25	TTGGGCGATT GCACATATGA CAGGTGTTTT GAGTAAAGTT GAGAGTTTTT ACCTAAACGA	15240
	CAAACGCAAT GAATCGATGC TCCATTATAC ACGCACAGAG AAGATTAAAC AGATGTATAA	15300
	AGAAATAGAT GTAAATGAGA TGGAAAGTCT TGTAGGCGCT AAGTTTGTA AATTATTTAC	15360
30	AGATATTGAT TTGAATGATG ATGAAGTCAT TTCAATATTT GTTTTCGATA AGTCAATAGA	15420
	ATAAGTGTTG CTGGTGTAAG GTACACGGTG CTGTTTGCTA ACTTCGCTTT GAATTTAACA	15480
	ATAATTCAAG GGGGTGGTAT GTCAAACGGT GCCGTTTTTT TGTCATATTT TTAAAACAAG	15540
35	CAACATGCAA CACGTACTTT AAGGAAGTCA AAATTTATCA TTTAGGAGAG ATGGATATGA	15600
	AAATCGTAGC ATTATTTCCA GAAGCAGTAG AAGGTCAAGA AAATCAATTA CTTAATACTA	15660
40	AAAAGCATT AGGATTAAAA ACATTTTATAG AGGAAAGAGG ACATGAGTTC ATTATATTAG	15720
	CAGATAATGG TGAAGACTTA GATAAACATT TACCAGATAT GGATGTGATT ATTAGTGCGC	15780
	CATTTTATCC TGCATATATG ACTCGTGAAC GTATTGAAAA AGCACCGAAC TTGAAATTAG	15840
45	CAATTACAGC AGGTGTAGGA TCTGACCATG TAGATTTAGC GGCAGCAAGT GAACACAATA	15900
	TTGGTGTGCT TGAAGTTACA GGAAGTAATA CAGTTAGTGT GGCAGAACAT GCGGTTATGG	15960
	ATTTATTAAT ACTTCTTAGA AACTATGAAG AAGGTCATCG TCAATCAGTA GAAGGTGAAT	16020
50	GGAAC TTGTC TCAAGTAGGT AATCATGCGC ATGAATTACA ACACAAAACA ATTGGTATTT	16080

55

TACAACACTA TGATCCAATC AATCAACAAG ACCATAAAATT GTCTAAATTT GTAAGCTTTG 16200  
 ATGAACTTGT TTCAACAAGT GATGCGATTA CAATTCATGC ACCATTAACA CCAGAAACTG 16260  
 5 ATAACCTTATT TGATAAAGAT GTTTTAAAGTC GTATGAAAAA ACACAGTTAT TTAGTGAATA 16320  
 CTGCACGTGG TAAAATTGTA AATCGCGATG CGTTAGTTGA AGCGTTAgCA TCCGAGCATT 16380  
 TACAAGGATA TGCTGGTGAT GTTTGGTATC CaCAACCTGC ACCTGCTGAT CATCCATGGA 16440  
 10 GAACAATGCC TAGAAATGCT ATGACGGTTC ACTATTCAGG TATGACTTTA GAAGCACAAA 16500  
 AACGTATTGA AGATGGAGTT AAAGATATTT TAGAGCGTTT CTTCAATCAT GAACCTTTCC 16560  
 AAGATAAAGA TATTATTGTT GCAAGTGGTC GTATTGCTAG TAAAAGTTAT ACAGCTAAAT 16620  
 15 AGAATAAGGA TGCTGGGCTA GCGATTAAACG CTTTCAATTT TATATAAATG AATCATATAA 16680  
 GCACTACTGC TGTGTAAAG ATGGCAGTAG TTTTTTTATG ATTACATCTA AGTATAGTCA 16740  
 CGGCTATGTT AGGACAATGA TTTAACATTT ACGCACATAT GTGTTCACTT ACGCAATTAT 16800  
 20 TGAnAAATnT CATTTCATGTG GnAATC 16826

(2) INFORMATION FOR SEQ ID NO: 47:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4012 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TTCAATGAGA GTAGTGGGCT GATGTTTAGC GATATCGCGT AAGATTAAAC ATTGGCCATA 60  
 35 ATATATATTG TGTTTTTCTA AAATCGGCTC GGCTAATTTT AAATAGGGGC GATATATTGT 120  
 TAT~~AAAA~~ACTA TTGAAAAATT CTTGTGATAG CATAGTGACA TCTCCTAAGA CAAAATAGTT 180  
 AGCTTAGCTA mCCTTTTTTAC AACAATAGTA ATTATAAAAC GGGAGCAATT AGAAATCAAT 240  
 40 ATATAATTAT TAAGAGCAAA AATAATTATA CTTTGTTAAA ATAAGCGTAA TTACATGTAA 300  
 ATAGGGGGAT ACTAATGATA TTGAAATTG aTCACATCAT TCATTATATA GATCAGTTAG 360  
 ATCGGTTTAG TTTTCAGGA GATGTTATAA AATTACATTC AGGTGGGTAT CATCATAAAT 420  
 45 ATGGAACATT CAATAAATTA GGTATATCA ATGAAAATTA TATTGAGCTA CTAGATGTAG 480  
 AAAATAATGA AAAGTTGAAA AAGATGGCAA AAACGATAGA mGGCGGAGTC GCTTTTGCTA 540  
 CTCAAATTGT TCAAGAGAAG TATGAGCAAG GCTTTAAAAA TATTTGTTTG CGTACAAATG 600  
 50 ATATAGAGGC AGTTAAAAAT AACTACAAA GTGAGCAGGT TGAAGTAGTA GGGCCGATTC 660

55

EP 0 786 519 A2

	ATCAGGATGA TGATGAAATT AAGCCACCAT TTTTATTCA ATGGGAAGAA AGTGATTCCA	780
	TGCGTACTAA AAAATTGCAA AAATATTTTC AAAACAATT TTCAATTGAA ACTGTTATTG	840
5	TGAAAAGTAA AAACCGATCA CAAACAGTAT CGAATTGGTT GAAATGGTTT GATATGGACA	900
	TTGTAGAAGA GAATGACCAT TACACAGATT TGATTTTAAA AAATGATGAT ATTTATTTTA	960
	GAATTGAAGA TGGTAAAGTT TCAAAATATC ATTCGGTTAT CATAAAAGAC GCACAAGCAA	1020
10	CTTCACCATA TTCAATTTTT ATCAGAGGTG CTATTTATCG CTTTGAACCA TTAGTATAAA	1080
	TATACGTAAG TGCTATGAGC GAGAATGCCC ATATGAATAA TGACAAGCAC AATGGAAAGA	1140
	ATCGTTAATA TATTATTTAA TCGTGATGAC TTAATTAAAA TGAAAAAGAT TGATAATATA	1200
15	AATGTGAAAA AGATAAGTAT AACCCGTAAA CTAAGTAAT TCACGGTGAG AGGTTGACTC	1260
	AATGTCATAA TGATTGCAAC GATGTTTATA ATTATAAATA GACTTAAAAT AATTGTTCTC	1320
	ATATCAAACA CCTCATTGTT AGATTATTGA CATTATAACA GGGGTAATTG TATATGAACA	1380
20	TTAATGTGGT TGCTTGAGGA AAAATTTATT CATTGAAGTC AAGTTGGTTC ATTTTAGAAA	1440
	TGAATATCGT GTTAGATGAT GAAAGTATAT TGAAGTATAG GTAAGTAGTT GAAAAGTATT	1500
25	AATTGTACGA TAACATTAAA TTTAACACGA AACATAGATA TAAATGATT CACAATTAAA	1560
	ATGGGTAAAT TTGAACTTGC TAAACTATTA ATTGGAGCAT GGACATTTCA AAAATAAGAG	1620
	TTCAAATCTT ACACAAGCTC TGAATCGACA CTATAAGATA CAACTGTAT AATTAAAGGT	1680
30	ATTGTAAAT AGAAGGAGAT ATCATAAATC ATGGAAGA TGATATCAC TAATCAGGAA	1740
	CATGACGCAT TTGTTAAATC CCACCAAAT GGAGATTTAT TACAATTAAC GAAATGGGCA	1800
	GAAACAAAGA AATTAAGTGG ATGGTACGCG CGAAGAATCG CTGTAGGTCG TGACGGTGAA	1860
35	GTTCAGGGTG TTGCGCAGTT ACTTTTAAA AAAGTACCTA AATTACCTTA TACGCTATGT	1920
	TATATTTTCG GTGGTTTGT TGTGATTAT AGTAATAAAG AAGCGTTAAA TGCATTGTTA	1980
	GACAGTGCAA AAGAAATTGC TAAAGCTGAG AAAGCGTATG CAATTAAAAT CGATCCTGAT	2040
40	GTTGAAGTTG ATAAAGGTAC AGATGCTTTG CAAAATTGA AAGCGCTTGG TTTTAAACAT	2100
	AAAGGATTTA AAGAAGGTTT ATCAAAAGAC TACATCCAAC CACGTATGAC TATGATTACA	2160
	CCAATTGATA AAAATGATGA TGAGTTATTA AATAGTTTGG AACGCCGAAA TCGTTCAAAA	2220
45	GTGCGCTTGG CTTTAAAGCG AGGTACGACA GTAGAACGAT CTGATAGAGA AGGTTTAAAA	2280
	ACATTTGCTG AGTTAATGAA AATCACTGGG GAACGCGATG GCTTCTTAAC GCGTGATATT	2340
	AGTTACTTTG AAAATATTTA TGATGCGTTG CATGAAGATG GAGATGCTGA ACTATTTTAA	2400
50	GTAAAGTTGG ATCCAAAAGA AAATATAGCG AAAGTAAATC AAGAATTGAA TGAACCTCAT	2460

55



CAAAATATGA TTAATGATGC GCAAAATAAA ATTGCTAAAA ATGAAGATTT AAAACGAGAC 2580  
 CTAGAAGCTT TAGAAAAGGA ACATCCTGAA GGTATTTATC TTTCTGGTGC ACTATTAATG 2640  
 5 TTTGCTGGCT CAAAATCATA TTAATGATGC GGTGCGTCTT CTAATGAATT TAGAGATTTT 2700  
 TTACCAAATC ATCATATGCA GTATACGATG ATGAAGTATG CACGTGAACA TGGTGCAACA 2760  
 ACTTACGATT TCGGTGGTAC AGATAATGAT CCAGATAAAG ACTCAGAACA TTATGGATTA 2820  
 10 TGGGCATTTA AAAAAGTGTG GGGAACATAC TTAAGTGAAA AGATTGGTGA ATTTGATTAT 2880  
 GTATTGAATC AGCCATTGTA CCAATTAATT GAGCAAGTTA AACCGCGTTT AACAAAAGCT 2940  
 AAAATTAATA TATCTCGTAA ATTAATAACGA AAATAGATTA ACGACTGAAA TCTGAACGCT 3000  
 15 CATAAGACTG TCATTTGCGT TCAGATTTTT TTACACAATA TAGAATGGTT GAGTAAAATA 3060  
 TTTTGAATA TAGTGAAAGA GGGGGAAGTA CTGTGATAAA AAAGCTATTA CAATTTTCTT 3120  
 TAGGGAATAA GTTTGCTATC TTTTAAATGG TTGTTTTAGT TGTCTTGGGC GGTGTATATG 3180  
 20 CGAGTGCTAA ATTGAAATTA GAATTACTAC CAAATGTACA AAATCCAGTT ATTTCAAGTTA 3240  
 CAACAACAAT GCGGGGTGCA ACGCCACAAA GTACCCAAGA TGAAATAAGT AGTAAAATTG 3300  
 25 ACAATCAAGT AAGATCATTG GCATATGTGA AAAATGTTAA AACGCAATCC ATACAAAATG 3360  
 CTTCAATTGT AACAGTTGAA TATGAAAATA ATACAGATAT GGATAAAGCA GAAGAACAGC 3420  
 TTAAAAAGA AATCGATAAA ATTAATTTA AAGATGAAGT TGGTCAACCA GAATTAAGAC 3480  
 30 GTAATTCGAT GGATGCTTTT CCGGTTTATG CATATTCATT TTCAAATAAA GAGAATGACT 3540  
 TGAAAAAAGT AACGAAAGTA CTGAATGAAC AATTAATACC AAAATTGCAA ACGGTAGATG 3600  
 GTGTGCAAAA TGCAGCAATTA AATGGGCAGA CGAACCGTGA AATCACCCTT AAATTTAAGC 3660  
 35 AAAATGAACT TGAAAAATAT GGGTTGACTG CTGATGATGT AGAAACTAT CTAAAAACGG 3720  
 CAACAGAAGC AACGCCACTT GGATTGTTCC AATTGTTGTA TAAAGATAAT CAATTGTTGT 3780  
 TGATGGTCAA TATCAATCTG TTGATGCTTT TAAAAACATA AATATTCCAT TAACGTGGCA 3840  
 40 GGAGGACCAA GGCATCTCA TCCCAAAGTG ACCATAAACC AAATTCAGCC ATGTCAGACG 3900  
 TTATCAGGCA TCACCACAGC AAATCAAAG CGTCAGCnCC AATATATAGT GGATGCCGCA 3960  
 nGAACTAGGG GTTTAGCGnT ATCAGTGGTG TGGCGACTCT ATTCTAAACG AT 4012  
 45

(2) INFORMATION FOR SEQ ID NO: 48:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7778 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

	CAATATAGGT CGCCGAGTTT CAACTaCATC AACTGGTTCA GTTACATTAG ATAATGCGCT	60
5	AGGTGTAGGT GGCTATCCTA AAGGACGAAT TATTGAAATT TATGGTCCTG AAAGTTCTGG	120
	TAAGACAACA GTAGCGCTTC ACGCTATTGC TGAAGTACAA AGTAATGGCG GGGTGGCAGC	180
	ATTTATCGAT GCTGAACATG CTTTAGATCC AGAATATGCT CAAGCATTAG GCGTAGATAT	240
10	CGATAATTTA TATTTATCGC AACCGGATCA TGGTGAACAA GGTCTTGAAA TCGCCGAAGC	300
	ATTTGTTAGA AGTGGTGCAG TTGATATTGT AGTTGTAGAC TCAGTTGCTG CTTTAACACC	360
	TAAAGCTGAA ATTGAAGGAG AAATGGGAGA CACTCACGTT GGTTTACAAG CTCGTTTAAAT	420
15	GTCACAAGCG TTACGTAAAC TTTCAGGTGC TATTTCTAAA TCAAATACAA CTGCTATTTT	480
	CATCAACCAA ATTCGTGAAA AAGTTGGTGT TATGTTCCGT AATCCAGAGA CTACACCAGG	540
20	TGGACGTGCA TTAATAATTCT ATAGTTCAGT AAGACTAGAA GTACGTCGTG CAGAACAGCT	600
	TAAACAAGGA CAAGAAATTG TAGGTAATAG AACTAAAATT AAAGTCGTTA AAAATAAAGT	660
	GGCACCACCA TTTAGAGTAG CTGAAGTTGA TATTATGTAT GGACAAGGTA TTTCTAAAGA	720
25	GGGTGAACTT ATTGATTTAG GTGTTGAAAA CGACATCGTT GaTAAATCAG GAGCATGGTA	780
	TTCTTACAAT GGCGAACGAA TGGGTCAAGG TAAGGAAAAT GTTAAAATGT ACTTGAAAGA	840
	AAATCCACAA ATTAAAGAAG AAATTGATCG TAAATTGAGA GAAAAATTAG GTATATCTGA	900
30	TGGTGATGTT GAAGAAACAG AAGATGCACC AAAGTCATTA TTTGACGAAG AATAGTACAC	960
	AAATTTATAT CTATAGTTAA ACTTAGCAAA TATCCTTATA GGATTGATTG AAAGTGATAT	1020
	TCATCTCATA AAGCTAGAAT AATATCTAAC TTTATGGGAT AACTACAAA TCGAGACTAT	1080
35	AAGGTTTTTT ATTTTATTTA TTATTACATT ATCAATAGTT TTATAATCGA GCTTCAAAAC	1140
	TTTAGAAAAT AGTAGAAATA GCATTCAATA TAGTGCAAAA GTGCAAATTG ATAACCTGAC	1200
	ACTTATCTCC TATAAACCGT ACAATTAATT TGTATGATTT ATATATAATT TCATAAAGTC	1260
40	ATATTGAATT TCATATAAAG AGCAAACCTT AGAAAAGGAG GTGTTTGTGT GAATTTATTA	1320
	AGCCTCCTAC TCATTTTGCT GGGGATCATT CTAGGAGTTG TTGGAGGGTA TGTGTTGCC	1380
	CGAAATTTGT TGCTTCAAAA GCAATCACAA GCTAGACAAA CTGCCGAAGA TATTGTAAAT	1440
45	CAAGCACATA AAGAAGCTGA CAATATCAAA AAAGAGAAAT TACTTGAGGC AAAAGAAGAA	1500
	AACCAAATCC TAAGAGAACA AACTGAAGCA GAACTACGAG AAAGACGTAG CGAACTTCAA	1560
50	AGACAAGAAA CCCGACTTCT TCAAAAAGAA GAAAACTTAG AGCGCAAATC TGATCTATTA	1620
	GATAAAAAAG ATGAGATTTT AGAGCAAAAA GAATCAAAAA TTGAAGAAAA ACAACAACAA	1680

55

	CGCATCTCCG GTCTCACTCA AGAAGAAGCT ATTAATGAGC AACTTCAAAG AGTAGAGGAA	1800
	GAACTGTCAC AAGATATTGC AGTACTTGTT AAAGAAAAAG AAAAAGAAGC TAAAGAAAAA	1860
5	GTTGATAAAA CAGCAAAAGA ATTATTAGCT ACAGCAGTAC AAAGATTAGC AGCAGATCAC	1920
	ACAAGTGAAT CAACGGTATC AGTAGTTAAC TTACCTAATG ATGAGATGAA AGGTGGAATC	1980
	ATTGGACGAG AAGGACGAAA CATCCGCACA CTTGAAACTT TAACTGGCAT TGATTTAATT	2040
10	ATTGATGACA CACCAGAAGC GGTATATTA TCTGGTTTTG ATCCAATAAG AAGAGAAATT	2100
	GCTAGAACAG CACTTGTTAA CTTAGTATCT GATGGACGTA TTCATCCAGG TAGAATTGAA	2160
	GATATGGTCG AAAAAGCTAG AAAAGAAGTA GACGATATTA TTAGAGAAGC AGGTGAACAA	2220
15	GCTACATTTG AAGTGAACGC ACATAATATG CATCCTGACT TAGTAAAAAT TGTAGGGCGT	2280
	TTAAACTATC GTACGAGTTA CGGTCAAAAT GTACTTAAAC ATTCAATTGA AGTTGCGCAT	2340
	CTTGCTAGTA TGTTAGCTGC TGAGCTAGGC GAAGATGAGA CATTAGCGAA ACGAGCTGGA	2400
20	CTTTTACATG ATGTTGGTAA AGCAATTGAT CATGAAGTAG AAGGTAGTCA TGTTGAAATC	2460
	GGTGTAGAAT TAGCGAAAAA ATATGGTGAA AATGAAACAG TTATTAATGC AATCCATTCT	2520
25	CATCATGGTG ATGTTGAACC TACATCTATT ATATCTATCC TTGTTGCTGC TGCAGATGCA	2580
	TTGTCTGCGG CTCGTCCAGG TGCAAGAAAA GAAACATTAG AGAATTATAT TCGTCGATTA	2640
	GAACGTTTAG AAACGTTATC AGAAAGTTAT GATGGTGTAG AAAAAGCATT TGCGATTGAG	2700
30	GCAGGTAGAG AAATCCGAGT GATTGTATCT CCTGAAGAAA TTGATGATTT AAAATCTTAT	2760
	CGATTGGCTA GAGATATTAA AAATCAGATT GAAGATGAAT TACAATATCC TGGTCATATC	2820
	AAGGTGACAG TTGTTGAGAG GACTAGAGCA GTAGAATATG CGAAATAATT TTTGTCTCCC	2880
35	TCACAAATTA GTGAGGGAGC TTTTTTAAGT TGTAAGTCTTA ATCTAGTTAG ACAGCACTTT	2940
	ATCGGTAATA ACTATATTAA ACAGTAGTTA TTTGAAAGTA AGACGGACCT TATATTAAAT	3000
	AAGAAGTTAT TGCTTTTAAT AAAAATGTTT TAGGCTTCGT AATTACTATA TTTATATTAT	3060
40	GTAAACCTAT AAAGATGATT GGTTTTCTAT CCAATAAAAA AGAAGAGAAG ATGTAACACA	3120
	TCTTCTCTTC YGCAATATTA ATTAGGATTT ATTTCTAAGT TGAGTTATTT TAATTGTAAA	3180
	TCTGTTTTCT TTAATTCCTT TATAACTTCT GCAGTATCAT AACAAATTGT TGCAATTGTT	3240
45	GAATATCTCT CTGCTAAACG ATATGCATTA ATGTAAAGCT TTAAACTTTC TTTAGCTATA	3300
	TCCTCTGCAT CTTGGAATTT TGATGGGTTA GACATAACCA CTAATTCTGC AAATTTTTCT	3360
50	GGATCAATAT TAATAGACAT GTATTTATTT ACAACTCCTA TTTATTTTGA TGTCTTAATA	3420
	CTAACATATT GAAGTTTTCA GACAAAGTAA TGTCTCTCTA TAATTGAAGA AAAATAATTC	3480

55

	GGATGAACAA AACATGAGAA TAATGTTTAT AGGGGATATC GTAGGTAAAA TTGGACGAGA	3600
	CGCAATTGAA ACGTACATAC CTCAACTGAA GCAAAAGTAT AAACCAACAG TTACAATTGT	3660
5	AAATGCTGAA AATGCAGCAC ATGGTAAAGG TTTGACTGAA AAAATATATA AACAATTACT	3720
	AAGAAATGGT GTAGATTTC A T GACTATGGG TAATCACACA TATGGTCAAC GTGAAATTTA	3780
	TGATTTTATA GATGAAGCAA AACGACTAGT AAGACCAGCG AATTTTCCGG ATGAAGCGCC	3840
10	GGGAATTGGT ATGAGATTTA TACAAATTAA TGATATTAAA CTTGCAGTTA TTAATCTGCA	3900
	AGGAAGAGCG TTTATGCCAG ATATTGATGA TCCTTTTAAA AAGGCAGATC AATTAGTCAA	3960
	GGAAGCACAA GAACAACTC CGTTTATATT TGTTGATTTT CATGCAGAAA CAACTTCTGA	4020
15	AAAGTATGCA ATGGGATGGC ATTTAGATGG TAGAaTAGCG CTGTTGTTGG AACGCATACA	4080
	CACATTCAAA CAGCAGATGA ACGTATTTTA CCAAAGGGGA CAGGGTATAT AACGGATGTT	4140
	GGTATGACAG GTTTTTATGA TGGCATTTTA GGAATAAATA AAACAGAGGT AATTGAGCGT	4200
20	TTTATCACTA GTTTGCCACA AAGACATGTT GTTCCAAATG AAGGTAGAAG TGTATTATCT	4260
	GGTGTGTGTA TTGATTTAGA CAAAGAAGGT AAAACAAAGC ACATCGAACG TATATTGATA	4320
25	AATGATGACC ATCCATTTTC AACATTTTAA AATTACGTAA GTAAACATTC GAATTGGACC	4380
	CTATCGTCCA TTAGTATGAA TTTAATATAG TACCACTGTT TACATAGTAA ATCGGTGGTT	4440
	CTTTTGTGTA TCATTTAATA TGAAATATAT CCATAGGAGG CATATACTA TGAAACCACA	4500
30	ATTATCGTGG AAAGTTGGCG GTCAACAAGG CGAAGGTATT GAATCAACTG GGGAAATCTT	4560
	CGCTACGGCT ATGAATAGAA AAGGATATT A TTTATATGGA TATAGACATT TTTCAAGTCG	4620
	TATCAAAGGT GGACATACGA ATAATAAAAT TAGAGTTTCT ACGACGCCTG TTCATGCAAT	4680
35	TAGTGATGAT TTAGATATTT TGATTGCATT TGACCAAGAA ACAATTGATG TTAACCATCA	4740
	TGAAATGAGA GAAGACAGTA TTATTTTAaC TGATGCCAAG GCTAAACCTG TGAAaCCAGA	4800
	AGGATGTCAT GCACAGCTTA TTGAATTACC TTTTACAGCA ACCGCTAAAG AATTAGGTAC	4860
40	AGCATTAAATG AAAAACATGG TTGCAATAGG TGCTACTAGC GCATTGATGA ATTTGAATAC	4920
	AAATACATTT GAAGAACTTA T TACTAATAT GTTTTCTAAA AAAGGTGACA AGGTAGTTGA	4980
	AGTCAATATC CAAGCATTAA ACGAAGGTTA TCAATTAATG CAATCTCGCT TACCTGAAAT	5040
45	CTACGGGGAC TTTGAATTAG AGTCAACAGA TGCACTACCA CATCTATATA TGATTGGTAA	5100
	CGATGCCATT GGATTAGGTG CAATTGCTGC AGGTTCAAA TTTATGGCGG CATATCCTAT	5160
50	TACACCTGCG TCTGAAGTTA TGGAATATAT GATTGCCAAT ATATCTAAAG TAAACGGAGC	5220
	GGTTATTCAA ACAGAAGATG AAATTGCTGC TGTAACATG GCTATTGGTG CAAATTATGG	5280

55

	TGGATTATCT	GGTATGACTG	AAACGCCATT	AGTCATTATT	AATACCCAAC	GAGGTGGACC	5400
5	TTCTACTGGA	TTACCTACGA	AACAAGAAC	GTCAGATTTA	ATGCAAATGA	TTTATGGTAC	5460
	ACATGGTGAT	ATTCCAAAA	TTGTTGTAGC	ACCAACAGAT	GCAGAAGATG	CATTTTATTT	5520
	AACTATGGAA	GCATTTAATT	TAGCAGAACA	ATATCAATGC	CCTGTTATAG	TTCTAAGTGA	5580
10	TTTGCAATTA	TCTTTAGGTA	AACAACTGT	TGAAAAATTA	GATTATAATC	GTATTGAAAT	5640
	TAAACGTGGT	GAAATCATTC	AATCTGATAT	TGAACGTGAA	GAAGATGATA	AAGGTTATTT	5700
	CAAGCGTTAT	GCGTtAACAT	CCGATGGTGT	TTCTCCTAGA	CCTATCCCCG	GTGTtAAAGG	5760
15	AGGTATTCAT	CATATAACTG	GTGTGGAAaCa	CAATGAAGAA	GGTAAACCTA	GTGAATCTGC	5820
	GTCAAATAGA	CAACAACAAA	TGAAAAACG	AATGCGTAAA	ATTGAGCAGT	TACTAATTGA	5880
	ATCGCCAGTA	GAAGCTAACT	TACAACATGA	GGATGCAGAT	ATTCTTTATA	TCGGTTTTAT	5940
20	TTCTACAAAA	GGTGCAATTC	AAGAAGGTAG	TAACCGTTTG	AATCAACAAG	GCATAAAAGT	6000
	TAACACTATA	CAAATTAGAC	AATTGCATCC	ATTCCCAACA	AGCGTTATTC	AAGATGCAGT	6060
25	TAATAAAGCG	AAGAAAGTCG	TTGTAGTGA	GCACAATTAT	CAAGGACAAT	TGGCTAGTAT	6120
	TATAAAAAATG	AATGTCAATA	TTCATGATAA	GATTGAAAAT	TATACAAAGT	ATGATGGGAC	6180
	ACCTTTCTTA	CCACATGAAA	TCGAAGAAAA	AGGCAAAATA	ATTGCTACTG	AAATAAAGGA	6240
30	GATGGTATAG	ATGGCGACAT	TTAAAGATTT	TAGAAATAAT	GTTAAGCCTA	ACTGGTGCCC	6300
	CGGATGTGGC	GATTTCTCAG	TACAAGCTGC	AATTCAAAAA	GCAGCCGCAA	ATATAGGGTT	6360
	AGAACCTGAA	GAAGTAGCTA	TCATCACC GG	TATAGGATGT	TCTGGCCGTC	TTTCAGGATA	6420
35	TATTAATTCT	TATGGCGTTC	ATTCTATTCA	CGGACGTGCA	TTACCTTTAG	CTCAAGGTGT	6480
	AAAAATGGCG	AATAAAGATT	TAACTGTTAT	TGCATCGGGA	GGAGATGGTG	ATGGTTATGC	6540
	TATAGGTATG	GGGCATACAA	TCCATGCTTT	AAGAAGAAAT	ATGAACATGA	CGTATATAGT	6600
40	CATGGATAAT	CAAATTTATG	GTTTGACAAA	GGGACAAACA	TCGCCGTCAT	CAGCAGTAGG	6660
	ATTTGTTACT	AAAACAACGC	CAAAAGGTAA	TATAGAAAAA	AATGTTGCGC	CTTTAGAATT	6720
45	AGTATTATCA	TCTGGTGCCA	CATTTGTAGC	CCAAGGTTTT	TCAAGCGATA	TTAAAGGATT	6780
	AACAAAACATA	ATTGAAGATG	CAATTAATCA	TGATGGATTT	TCATTCGTTA	ATGTCTTTTC	6840
	ACCATGTGTG	ACTTATAATA	AAATTAACAC	ATACGATTGG	TTTaAAGAAC	ATTTAACAAG	6900
50	TGTTGATGAc	ATTGAAAATT	ATGATTCTAC	AGATAAACAA	TTAGCGACTA	AAACTGTTAT	6960
	TGAACATGAA	TCTTTAGTAA	CTGGTATTGT	TTATCaAGAT	AAAGAAACAC	CATCATATGA	7020
55	ATCtCAAATT	AAAGAGTTAG	ATGATmCACC	ACTTGCTAAA	AGAGATATCa	AAATTaCTGA	7080

TGTATTTATA ACAGATCCAT TTATGCTACT CAGTTTTTTA CTATTACAAA AAATAAAGGA 7200  
 GTTTTTAAAA ATGAAAGACA CATTAAATGAG TATACAAATA ATTCCTAAAA CACCAAACAA 7260  
 5 TGACAATGTT ATACCTTACG TAGACGAGGC GATTAAAAATA ATTGACGAAT CTGGTTTGCA 7320  
 TTTTAGAGTA GGTCCGTTAG AAACGACAGT ACAAGGAAAT ATGAATGAAT GTTTAATTTT 7380  
 10 AATACAATCA TTAAATGAAC GAATGGTGGA ACTTGAATGT CCAAGTATTA TTAGCCAAGT 7440  
 TAAGTTTTAT CATGTGCCAG ATGGCATCAC TATTGAACT TTAAGTAAA AATATGATGA 7500  
 ATAACATTAA AAGTGAAGTA AACTGGATTT GAATTGGCTT GTTAGAGATG ACGTATAACT 7560  
 15 TTAAGTGT TTGCACTTTA TAGTTAAATT TAATATAATT ATTAAATGAT ACGGGCAAAT 7620  
 AGAAAGGATT TTGTAAAGTG AACGAAGAAC AAAGAAAAGC AAGTTCTGTA GATGTTTTAG 7680  
 CTGAGAGAGA TAAGAAAGCA GAAAAAGATT ATAGTAAATA TTTTGAACAT GTTTATCAGC 7740  
 20 CGCCTAATTT AAAAGCAAGC GCAAAAAAG AGGThAAA 7778

## (2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 1128 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AGATGAAGTT GTTACgAAAA TTGCGTACGC TGTTTCAGAA CATGTCAAAA TAGAAACAGG 60  
 35 TAATCCATTc TTTCAAACAT CACATAGTGG TTGTGCGACG GGCGGATCCT GTAATTGTTC 120  
 ATTATAAAAA ACATCGAGTC AGAAAAAGGT GGTTATTGAA cCACTAACTA GCATCTGACT 180  
 40 CGATGTTTTT ATTTATTcGG GATTGTTTGT TTGAATTGTT GTGCTAAATC TGGTCGATCT 240  
 GTCACAATCG TGTGTGCACC TTTTGGTAT AAATCATTCa TCAGATTTAT ACTATTTACG 300  
 CCATAATAGC CTGGAATGAT ATTCATATCA TTTAACCATT TGATAAAACG AGATGAAGTC 360  
 45 AAATCAATGC CTTTAAAAATG AGTAGGCATT TGGAACGTTT GTGCTAATGG TTGGTAGTAC 420  
 CTACCACCTA ATAAATGATA TTTTAAAAAT GCTTCTGTAA CTTCTGTGTT GCTAGCACCA 480  
 ATTGCGACGG ATCCTTGTC AATTTTATTA AAACGAACGA TTTGTTCTTT ATAAAAACTT 540  
 50 GTCACAAGAA CGCGGTCAAA TGCTTGATTT TCTGCAATTG TATCAAACAT AATTTGTGGT 600  
 GCGATTGAGC CTTCATAGGA TTCAGGAGCA TCTTTTAAAGT CTACGTTTAT ATACATATCA 660  
 55 GGATATTGCT TCAGCAACTc ATCGAAGGTT AGTATAGCTG TGTGTGCATG ACCACGATAT 720

AATGTATGGG CACTAACTTT TCCAGAGCCG TTCGTCGTTT TATCAACAGT TGCATCATGA 840  
 AAAACGATAA GCTGTTGATC TTTTGTGAGT CTCACATCTG TTTCAAAGCC ATCAACGCCT 900  
 5 AATTGTTTAG CATAGTCAAA TGCAAGTTGC GTTTGCTCTG GTCTTAAAGC CATACCACCG 960  
 CGATGCGCAA ATATATATGG TGCAATTGCCT TTGAAAAAAG CAGGGATGGT TTGCTTTTTTA 1020  
 10 GTAATCACTT TATTTTTATT GATCATTAAAT AGACTACTTA AAAATCCAGC ACCGACTAGT 1080  
 ACCGCATTTA AAATGTTTCT GTTTACnTTT TTCATAAAAA ATTCCTCC 1128

(2) INFORMATION FOR SEQ ID NO: 50:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6252 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

25 CAAGCAAACA ATCGTCGATA AAATTGCTAA AATAATAAAA GTAATTCGAA CTTTCATCAT 60  
 GATCATCCTT TGTTTATAGA GTCAATATAA GTATGGAATA TGTTAGGTAT ATAGTCAAAT 120  
 GCGTCAACTA ATGGGAATTT TGGCATAGAT AGAGAATTTA AGGCAATTAA AAAGGCATCA 180  
 30 AACAGTAATA TGCTGCTTGA TGCCCAAATG ATGACTTTAG CTAAATTGAT TAGTCACTTT 240  
 TAAAGATAAA GAATTGTCAT GAATTAAAAC TCATGTAATG ATGTGTTACA TTTCGCAATG 300  
 ATGGCTTTCA GTTATTTATC GATAACATCA CTCTTGATAC CTTTAGATTT TAAGAAATCT 360  
 35 TTAATTTTAT CTTGTTGCTT TTTATTAACA TCACCGGCAT ATTTTGTGG CACGTCGACA 420  
 ACATTGATTT TATTTTGCGG TTGATAGCTA AGCTTTTCAA TATCTTCATC AACATTGGCG 480  
 ATTGACTAT TTAAGCTTT GAAGTAATTC ATCATTAAAT CAACGGGTTT CTTATATTCT 540  
 40 TTAGGAATAT TGTTTTAGT GACAAATTTT TTGAAATGCA AATCGTTTTT AACAGCTAAG 600  
 TTAGATAAGT GGCTAAGTGT TTCTGCTTGT TTTTCAGTCA CTTTGTGTTG ACTGTCAATT 660  
 45 TGTTTATCTA GTTTATGTTG CATAATATAT TTGTTATCAA GTATATCGCT ATTTACAGAC 720  
 AAATACTTTT CTATAGCTTG CTTTCTCTCT GCATCACTAA TATCACTATT TTTCTTATCT 780  
 GAGTTAAAGA TATCTTTTGT TCTAATTTT TTAGCGCTTT TAGGTGCATG GATGCCAGTA 840  
 50 CTTGTATGAT GATCTTCGTT ATCAGATTGA TCGGACGCGC AACCTGTAAG AATTAATGTC 900  
 GATGCTAAAA ATGTACTTAG TAGTAATCTC TTTTTCATAA TGTAATATAA CTCCTTAGTT 960  
 55 TATCTTTAAT TGAAAAAATA TGTATTCATG TTAATAGAG TAACATTGAA TTAGTTTGGA 1020

	TCTATCAATA ATGCATCATT TTGGACGTTG TTAAGGATAG CTTTATCTAT AAATAACTGC	1140
	ATAATTGGTT GTACTAATTT AGACGTAGGT ATCGTACGTA AAAGCATAAT AATTTTCGTTT	1200
5	ACATACTTTT CTTTCTCAAT ATCATTTTTTC ATATTGATTT GTTTGCGAGA GGTACATACT	1260
	TTAAGCATT A TCGCACATCT CGTTGTATAT ATTAAGTTTA TCATAACATG ATTTTATGTC	1320
10	GGGATAAAAA AATAACAGCA TCTTAACAAA TGTAAGATAC TGTCAGTGAA ATGAATGAAA	1380
	CTTTAGTTTC TGaTAATATA GTCAAAGGCA TTTAATGCTG CATTTGCACC AGCGCCCAT	1440
	GAAATGATAA TTTGTTTGT CTTCTGATCT GTGACATCGC CAGCAGCAAA TATTCCAGGA	1500
15	ACATTCGTAT TATTGTTACG ATCAATCACA ATTTACCAC GTTCGTTTAA TTCAACAGCA	1560
	TCGTTTAACC ATGATGTGTT TGAAGTAAA CCAATTTGAA CAAAGATACC ATCTAAGTTA	1620
	AGTAGATGTT CTTGCGCGGT GTTCATGTCT TCGTAACGTA TACCTGTAAC ATGGTCTTCT	1680
20	CCGACAACTT CAGTAGTTTT GGCATTTGTT TTGATATCAA CATTTGATAA AGAACGTAAA	1740
	CGATCTTGTA ACACGTTGTC TGCTTTTAAT TCGCTAGCGA ATTCGAATAA TGTAACATGA	1800
25	TTAACGATAC CAGCAAGGTC AATTGCTGCT TCAACCCAG AGTTACCGCC ACCGATAACT	1860
	GCTACGTCTT TATTTTCAA TAGAGGTCCG TCACAGTGAG GGCAGAATGC AACACCTTTA	1920
	TTAATCAATT GCTCTTCACC TGAATGTTT AGCTTACGCC AACCTGCACC AGTAGCAATA	1980
30	ATGACTGTTT TACTTTCTAA GACAGCACCG TTTTCTAAG TAACTTTAAT TGCTTCGTCA	2040
	GTCTTTTCGA TATCTGTAGC ACGTATACCT GTCATTGCAT CAATGTCATA TTGATCAATG	2100
	TGCGCTGCTA AGTTAGAAGA AAATTCAGAA CCAGTTGTTT CTTTAACAGT AATGAAGTTC	2160
35	TCAATACCAG CAGTATCATT AACTTGGCCA CCGATACGAT CAGCAACTAT ACCAGTACGT	2220
	AAACCTTTAC GTGCTGTGTA AATCGCTGCA CTACCACTAG CAGGACCACC ACCAACGATT	2280
40	AAGACATCAT AAGGTTCTTT ATTTTCAAAC TCAGATGCAT CTGCCGTACT GCCTAGTTTC	2340
	GAAAGAATAT CTTGGATTGT CATACGACCA TTGCCAAATT CTTGCCATT TAAAAAGACA	2400
	GCAGGGACTG CCATGATGTT TTCAGATTCT TCACGGAACA CTGCACCATC AATCATAGAA	2460
45	TGCGTGATGT TAGGGTTGAT CACACTCATT AAGTTAAGTG CTTGAACGAC ATCAGGACAT	2520
	TTTGTACACG TTAAACTAAT GAATGTTTCA AAATGGAATG AACCTTCTAA TTTTAAATT	2580
	TGGTCAATGA TTGACTGTTT TTCTTTAGGT GCACGACCAC TAACCTGTAA AATTGCTAAA	2640
50	ACAAGTGAGT TAAACTCGTG ACCTAATGGA ATACCTGCAA ATGTTACACC TGTTTCTTCG	2700
	CCAGGACGAT TGA CTGAGAA ACTTGGTGTA CGTTTAAAG ATTTTTCAGA AAGAGATAGT	2760
55	CTAGGTGACA TATCAGTAAT TTCTGTCAAC AAATCTTTAA GTTCTTTGGA TTTATCATCT	2820



	TGTTGTTTTA AATCAGCATT AAGCATGGTT GTAATGCCTC CTTAGATTTT ACCTACTAAA	2940
5	TCTAAACCAG GTTGCAATGT TTTAGCGCCT TCTTCCCAT TTAGCTGGGCA TACTTCGCCA	3000
	GGGTTTTTAC GAACATATTG AGCTGCTTTG ATTTTGTGAG CTAATGTACT AGCGTCACGG	3060
	CCAATCCGT CAGCGTTAAT TTCAGATGCT TGTACAACAC CGTCTGGGTC GATAATGAAT	3120
10	GTACCACGTT GAGCTAAACC AGTAGCTTCA TCTAATACAT CAAAATTACG AGTGATTGTT	3180
	TGTGATGGGT CACCAATCAT AGTGTAAGTG ATTTTGCTAA TTGCATCTGA ATGGTCATGC	3240
	CATGCTTTGT GTACGAAGTG AGTATCAGTT GATACTGAGA ATACATTTAC GCCTAATTTT	3300
15	TGTAATTCTT CATATTGGTT TTGTAAGTCT TCTAATTCAG TTGGACAAAC GAATGAGAAG	3360
	TCAGCAGGAT AGAAGCATAC TACGCTCCAA GAACCTTTTA AATCTTCTTG TGTAACCTCT	3420
	TTAAATTGAT CTTTTTTTGG ATCGAAACTT TCGCGTGTA ATGGTAAGAT TTCTTTGTTA	3480
20	ATTAATGACA TAAATATCTT CCTCCTAAGA ATTTAAGTAT GAATTAGAAC TATCAATTGA	3540
	TTGCGCTTAA TTATAATAAT TCTAATCTCT TAGTTAGCAT TATTACATTT TGATCCAGAA	3600
25	TAGTCAACTG GATAACTTTG TAAAGTGAAT GATTACTTTT AAAATAAAGA AAGATAATAT	3660
	AAAGTGCTTT GATAATGGAT TTTGTAGTTG ATGATTTAAA AGGTTGTGTC TATATTTAAT	3720
	ATCTTGATTT TAATGTAAAA AATGTAAAAA AAGAAGATTT GTATTCTCAA CTAAGTCAAC	3780
30	CTTATTGATA ATGGTATGAG AATATTTGTT CGAGATGGAT GAAGGTAATG AGTGAGAAAC	3840
	TGGATTTTTA AAGTATGAGA CAATATTTTA AAAAGTTCAA TTATTAACTT ATAAGCAAAT	3900
	AATTGCTATA AAAAAGTTTG GACGTGTACA ATTGCAATAT GAAGATTTTA AATTAAATTGT	3960
35	AAAGTATCGA GGAGTGGGTA ACGTGTGAGA ACATGTATAT AATCTTGTGA AAAAGCATCA	4020
	TTCTGTTAGA AAATTTAAGA ATAAACCTTT AAGTGAAGAC GTTGTTAAGA AATTGGTAGA	4080
40	AGCTGGACAA AGCGCTTCGA CGTCAAGTTT CCTGCAAGCA TACTCAATTA TTGGTATCGA	4140
	CGATGAGAAG ATTAAAGAAA ATTTACGAGA AGTTTCTGGA CAACCTTATG TTGTAGAAAA	4200
	TGGCTATTTA TTCGTCTTTG TTATTGATT AATATCGTCAT CATTAGTTG ATCAACATGC	4260
45	TGAAACTGAT ATGGAATG CATATGGTTC AACGGAAGGT TTGCTAGTAG GTGCAATCGA	4320
	TGCAGCATT A GTTGCCGAAA ATATTGCGGT AACTGCTGAA GATATGGGGT ATGGCATTGT	4380
	CTTTTATAGGA TCATTAAGAA ATGATGTTGA ACGCGTTCCA GAAATTTTAG ACTTACCTGA	4440
50	CTATGTCTTC CCGGTATTTG GTATGGCAGT AGGGGAACCC GCAGATGACG AAAATGGTGC	4500
	AGCCAAGCCA CGCTTACCAT TTGACCATGT CTTCCATCAT AATAAGTATC ATGCTGATAA	4560
55	GGAAACACAG TATGCACAAA TGGCAGATTA CGACCAGACA ATCAGCGAGT ACTATGATCA	4620

	CAAAGCAAGA TTAGATATGT TAGAACAATT GCAAAAATCA GGCTTAATAC AGCGATAgCA	4740
	AGATACCAAA ATAACCCGCC CCCCTCTAGC TTAAAATGAT AAGTATAGCT AGAGGGGGCG	4800
5	GGTATTTCTT GCAATGAATT AGTGTGAAGT TAATGCAGCA TTATCATTTG AATCGAAAGT	4860
	ATCTTTATCC CAATGTTTAG TTAACCTGGC GGTACCTGTA CCAGCTAGCA TTGAATCGTT	4920
10	CACGTTTAAT GCTGTTCTAC CCATGTCAAT CAATGGTTCA ACGGAGATGA GCACGCCGGC	4980
	TAAAGCGACT GGCAAGTTTA ACGTTGACAA CACCAATATG GATGCAAATG TAGCCCCGCC	5040
	ACCGACGCCA GCAACGCCGA ATGAAC TAATCAGACA GCGATTAACG TTACAATAAA	5100
15	TTGTAAATCA ATTTCTACAT TAGCGACGGG TCGGACCATA ATTGCAAGCA TGGCAGGGTA	5160
	AATGCCTGCA CAACCATTTT GTCCAATCGA CAATCCAAAT GTCGCAGCGA AATTGGCAAT	5220
	ACCTTCTGGC ACGCCTAGAC GTCTTGT TTGTGTACA TTCAATGGTA AGGCACCCGC	5280
20	GCTTGAGCGT GATGTGAATG CAAAGATTAA TACTTCCAAA GTCTTTTAA CATAGCGAAT	5340
	TGGGCTAATA CCTAACAGGC TTAAAATAAT TAAGTGAATG ATATACATCG TAATTAATGC	5400
25	AGCGTACGAT GCGATTAAGA ATTTTCCTAA AGTCCAAATG GCGCCAAAGT CACTTGTCGA	5460
	TAATGTGTTG GCCATAATTG CTAATACACC GTATGGCGTT AAACGTAAGA CGAACGTCAC	5520
	AATCGCCATT ACTAGTGAAT AGATAGCGTC AATCGCAGCG TTAAGCAATT CACCATGATC	5580
30	AGGTGTGTTG CGTGTACGCG TAAATAAGCA AATCCTATAA ACGAAGCAAA TATCAGACA	5640
	GCAATCGTGG aAGTTGCAG TTGTCCaGTG AAATCTAAGA ATGGATTTT AGGCAATAAT	5700
	TCCAAAATTT GTTGTGGTAA CGTATGTGCT GTTAAATCTT TCGCTTGTTT AGCAATTCG	5760
35	CTTCCACGTG CTTGTTACAGC GTTACCAAGG TTAATTGTTG ATGCATCTAA ACCAAACACC	5820
	AAGGCATACA CAACACCAAC AATCGCAGCA ATGGTGACAG TGCCAATTAA AAAGATAAAA	5880
40	ATGAGACTAC CAATTTTAGC AAACTTTCT CCGATTTGAA TTTTAGTGAA TGCAGCTACA	5940
	ATAGAAATGA AAATTAAAGG CATAACAATC ATTTGCAACA ATGCAACGTA ACCTTGTCGG	6000
	ACAATGTTGA ACCAGTCACT TGTTGATGTA ATAACATTCG AATGTGTGCC ATAAATAAGA	6060
45	TGCAATAACA CACCGAATAC TATACCAATC CCTAAAGCTG TAAACACACG TTTCGCAAAA	6120
	GATATATGTT TCGGAGCCAT CATGTGCAAT ATTACGATGA AAATCACCAA TACAATAATA	6180
	TTAATCAGTG TAAGAAAAGC ATTCATGAAC GTCACCTCTT AAATTTTGA ATATAATTCC	6240
50	GACTAGTATG CT	6252

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6730 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

	ATCAAATCnC AAAATATTTA TTAATnAnAA GGGGATTATC CaTGtgAGAA ACAAAGTAAT	60
10	GCTCTTTTTT TACCTCTTGT GGGTTGAAAA aTGGATCATC AGAGATAGAC TTCTTCTTTT	120
	TCGAAGATGA CATTTGATAC TTTAATCTTC TAAAACCATA ACTTGTGCGA TCAAAAATGC	180
	CTTCTTGTA C AAGTAAATC AAAAATATGC TAATAAAAAT AATTAATGAA ACATAAAACA	240
15	ATATATTTAA ATATGTAATG ATAGTATGGC TATTAAAAAG CCATATAATA AACGTTAATA	300
	TTGGCGTTAT TAGTGCCATT CCAAGCCATT TTTTCAACAT TTGATCACTC CCACTTATAG	360
	AAAACCTCTTA CGCATAGTTT ACATTAAAAT CAGACATTGA GGAATGATTT TTTAATTTCT	420
20	TCAGCTTTAT TGAAATTCTA AAATCAATCA TTCTTCATTA GTTTAAAGCA AAAAAATATT	480
	GATATATAGT AAATATTGTA TATATAATAT TAGTTAAGAT TTCaGAAAAT TTTGAAGGGA	540
25	ATGGAAATTT AGAAATCGGA ATTTGTTAGA GGAGGGGATT AGATGGGGAA ATATATTTTC	600
	AAACGATTTA TTTATATGCT TATTTCTTTA TTTATTATTA TTACAATTAC ATTTTCTTA	660
	ATGAAATTAA TGCCAGGTTT GCCATTTAAC GATGCTAAAT TAAATGCTGA ACAAAAAGAA	720
30	ATTTTAAATG AAAAATATGG ATTAAATGAT CCTGtAGCTA CGCagTATTT ACATTATTTA	780
	AAAAATGTTG TTACAGGCGA TTTTGGTAAT TCATTCCAGT ATCATAATCA ACCTGTGTGG	840
	GATTTGATTA AACCGAGACT ACTACCTTCT TTTGAAATGG GTCTTACAGC AATGTTCaTC	900
35	GGTGTGATAC TGGGACTTAT TTTAGGTGTT GCAGCAGCTA CTAAACAAAA TTCTTGGGTT	960
	GACTATACAA CTACAGTTAT TTCAGTTATT GCAGTATCTG TACCATCTTT TGTACTTGCT	1020
40	GTACTTTTAC AATATGTATT TGCAGTTAAA TTAAGATGGT TCCCAGTAGC TGGATGGGAA	1080
	GGTTTTTCGA CCGCGGTATT ACCGTCACTT GCATTATCTG CAGCTGTTTT AGCAACTGTC	1140
	GCCAGATACA TAAGAGCAGA GATGATAGAG GTATTAAAGT CAGACTATAT TTTATTAGCG	1200
45	AGAGCTAAAG GTAATTCGAC AATGCGTGTA CTTTTTGGAC ATGCACTTAG AAATGCTTTA	1260
	ATTCCAATTA TTACAATTAT CGTTCCCATG TTAGCAAGTA TTTTAACAGG CACTTTAACA	1320
	ATTGAAAATA TTTTGGAGT TCCTGGATTA GGGGATCAAT TCGTACGTTT AATTACAACA	1380
50	AATGATTTCT CAGTAATCAT GGCAATCACA CTATTATTTA GCACACTGTT TATCGTTTCT	1440
	ATTTTTATTG TAGATATTTT GTACGGTGTG ATAGATCCAC GAATTCGTGT TCcAAGgAGG	1500
55	TAAAAAATAA TGGCTGAAAA TAAAAACAAT TTGTCGATTA ACGACGATCA TTCTAATGCA	1560

	TGAATCAGGA ACCTGAAATG CAACGAGAAA GCAAAAACTT TTGGCAAGAT GCTTGGGCTC	1680
5	AGTTAAACG AAATAAGTTA GCTGTTGTCG GTATGATAGG TTTAATTATC ATTGTAATAT	1740
	TTGCTTTTAT CGGTCCAGTT ATAAATAAAC ATGATTATGC TGAACAAAAT GTAGAACATA	1800
	GAAATCTTCC GGCAAAAATA CCTGTATTAG ACAAGTTCC ATTTTACCT TTTGATGGTA	1860
10	AAGATGCAGA TGGCAAGGAT GCTTATAAAG CAGCAAATGC TAAAGAAAAT TATTGGTTTG	1920
	GTAATGATCA GTTGGGTCGA GATTTATGGA CAAGAACATG GAAAGGTGCT CAAATTCAT	1980
	TGTTTATCGG TGTTGTTGCA GCGATGTTAG ATATTTTAT TGGTGTGTA TATGGTGCGA	2040
15	TTTCTGGATT CTTGGGTGGA CGTGTGATA CGATTATGCA ACGTATACTT GAAGTCATAG	2100
	CATCTATTCC GAATTTAATT GTCGTAATTT TATTTGTATT AATTTTGA CCAATCCATT	2160
	GGACAATTAT ATTGGCTATG TCTATCACAG GCTGGTTAGG CATGAGCAGA GTTGTACGTG	2220
20	GAGAATTTT AAAATTAAAA AATCAAGAGT TTGTCATGGC TTCGAAAACA TTGGGGGCTT	2280
	CAAAATTCAA ATTGATATTT AAGCATATTT TACCTAATAC ATTAGGTGCT ATCGTGGTTA	2340
25	CATCAATGTT TACAGTACCT AGTGCTATTT TCTTCGAAGC ATTTTAAAGT TTCATTGGTA	2400
	TAGGTGTACC CGCACCTCAA ACATCGTTAG GGTCAATAGT AAATGATGGG CGCGCAATGT	2460
	TATTAATTTA TCCACATGAA TTATTTATAC CAGCAATGAT TTTAAGTTA TTAATTCTAT	2520
30	TCTTTTACTT ATTTAGTGAT GGATTACGTG ATGCATTGTA TCCGAAAATG CGTAAATAAA	2580
	AAGGGGGCAT AGCATATGAC TGAAAGAATA TTAGAAGTAA ATGATTTGCA TGTTTCCTTT	2640
	GATATTACAG CAGGGGAAGT GCAGGCAGTG AGAGGCGTAG ATTTTATTT GAACAAAGGG	2700
35	GAAACATTGG CAATTGTTGG TGAATCAGGT TCAGGTAAAT CTGTAACAAC AAAAGCAATT	2760
	ACAAAATTAT TCCAAGGGGA CACAGGAAGA ATTAATAAGG GAGAAATTTT ATTTTATGGG	2820
40	GAAGATTTAG CAAAAAACC TGAAATGAG TTGATTAAAT TACGTGGCAA AGATATTTCA	2880
	ATGATCTTTC AAGATCCAAT GACATCTTTA AACCCAACGA TGCAAATTGG TAAACAAGTC	2940
	ATGGAACCAT TAATTAAGCA CAAAATTAT AGTAAAGCAC AAGCTAAAAA GCGCGCATTG	3000
45	GAAATACTAA ATCTTGTAGG TTTACCAAAT GCAGAAAAAA GATTTAAAGC ATATCCTCAT	3060
	CAATTTTCAG GTGGACAAAG GCAAAGAATT GTTATTGCAA CCGCATTAGC TTGTGAACCT	3120
	AAAGTGCTCA TTGCTGATGA ACCAACGACT GCATTAGACG TAACGATGCA GGCACAAATT	3180
50	TTAGATTTAA TGAAAGAACT ACAACAAAAA ATCGATACAG CAATTATTTT TATAACGCAT	3240
	GATTTAGGGG TTGTTGCGAA TATTGCTGAT AGAGTGGCAG TTATGTATGG TGGTCAAATG	3300
55	GTTGAAACAG GAGATGTTAA CGAAATATTT TATGATCCAA AGCATCCATA TACATGGGGA	3360

	GGAGCGCCAC CTGATTTATT ACACCCACCT AAAGGTGATG CATTTCGAG ACGTAGcAAT	3480
5	ATGCATTAGA TATTGATTTT AAAGTAGAAC CACCGTGGTT TAAAGTTTCA CCGACACATT	3540
	TTGTGAAATC TTGGTTATTA GACGCACGTG CACCAAAAGT TGAAC TACCC GAGCTGGTAA	3600
	AACAACGTAT GAAACCGATG CCTAATAATT ATGAAAAACC ACTCAAGGTA GAAAGGGTGT	3660
10	CGTTCAATGA AAAATGATGA AGTGCTATTA TCTATTAAAA ATTTAAAGCA ATATTTTAAC	3720
	GCAGGAAAGA AAAACGAAGT GgaGCGATTG AAAATATTTT GTTTGATATA TACAAAGGGG	3780
	AAACATTAGG TTTAGTAGGA GAATCGGGGT GTGGTAAATC TACAACTGGT AAATCAATTA	3840
15	TTAAACTTAA TGATATTACA AGTGGAGAAA TTTTGTATGA GGGTATTGAT ATACAAAAGA	3900
	TTCGTAAACG TAAAGATTG CTTAATTTA ATAAAAAGAT ACAGATGATT TTTCAAGACC	3960
	CATATGCGTC TTTAAATCCT AGGTAAAAAG TAATGGATAT AGTAGCTGAA GGTATTGATA	4020
20	TCCATCATTT AGCAACTGaT AAGCGTGACC GAAAAAACG TGTCTATGaT TTACTTGaAA	4080
	CTGTTGGATT AAGTAAAGAA CATGCCAATC GCTATCCTCA TGAATTTTCA GGTGGaCAAC	4140
25	GCCAACGTAT TGGaATTGCC CGTGcATTAG CCGTTGaACC AGAATTCATT ATCGCGGACG	4200
	AACCAATATC GGCATTGGAT GTTTCAATCC AAGCTCAAGT AGTTAATTTA TTATTTAAAT	4260
	TACAACGTGA AAGAGGGATT ACGTTCCTAT TTATAGCTCA TGATCTATCA ATGGTGAAGT	4320
30	ATATTTT CAGA TCGTATTGCA GTCATGCATT TTGGGAAAAT AGTTGAAATT GGACCGGCAG	4380
	AAGAAATTTA TCAAAATCCA TTACACGATT ATACTAAGTC TTTATTATCA GCCATTCCAC	4440
	AACCTGATCC TGAATCAGAA CGCAGTCGCA AACGATTTAG TTATATTGAT GATGAAGCAA	4500
35	ATAATCATTT AAGACAATTA CATGAAATTA GACCGAATCA CTTTGTCTTT AGTACTGAAG	4560
	AAGAAGCGGC ACAACTACGA GAAAATAAAT TGGTGACACA AAATTAAGGG GAAGGGGGAA	4620
40	ATGcAATGAC GAGAAAATTT AGAACACTTA TTTTAATTTT GATTGCTACA ATTGCATTAA	4680
	GTGGTTGTGC TAATGACGAT GGTATTTTATT CAGATAAAGG TCAAGTATTC AGAAAAATTT	4740
	TGTCATCAGA CTTAACATCC CTTGATACAT CATTAATAAC GGATGAAATA TCTTCTGAAG	4800
45	TGAcTGCGCA AACATTGCGAA GGTTTATACA CATTAGGAAA AGGTGACAAA CCGGTGTTAG	4860
	GTGTTGCGAA AGCTTTTCCT GAAAAGAGTA AAGATGGTAA AACTTTAAAG GTTAAATTAA	4920
	GAAGCGATGC TAAATGGAGC AATGGTGACA AAGTGACTGC ACAAGACTTT GTTTATGCTT	4980
50	GGAGAAAAAC AGTTGACCCT AAAACAGGTT CTGAATTTGC ATACATTATG GGGGACATTA	5040
	AAAATGCGAG TGATATTAGT ACTGGTAAGA AACCTGTAGA GCAATTAGGT ATCAAAGCAT	5100
55	TAAATGATGA AACATTACAA ATTGAATTAG AAAAGCCGGT TCCATATATT AATCAATTAT	5160

ACGGTACGGC AGCTGATAGA GCGGTATACA ATGGTCCaTT TAAAGTTGAT GATTGGAAAC 5280  
 AAGAAGATAA AACCTTACTA TCTAAAAATC AGTATTATTG GGATAAAAAG AATGTAAAAT 5340  
 5 TAGATAAAGT GAATTATAAA GTTATTAAAG ACTTACAAGC CGGTGCAATCA TTGTATGATA 5400  
 CTGAATCAGT AGATGACGCA TTTATTACTG CAGATCAAGT AAATAAATAT AAAGACAACA 5460  
 10 AAGGATTAAA CTTTGTGTTA ACGACTGGGA CATTTTTTGT AAAAATGAAT GAAAAACAAT 5520  
 ATCCTGATTT TAAAAACAAA AATTTAAGAT TGsTATCGCA CAAGCAATAG ATAAAAAAGG 5580  
 ATACGTTGAT TCAGTGAAAA ACAATGGCTC AATTCCTTCC GATACACTAA CAGCCAAAGG 5640  
 15 AATTGCGAAA GCGCCTAATG GCAAAGATTA TGCGAGTACC ATGAATTCGC CTTTAAATA 5700  
 TAATCCTAAA GAAGCAAGAG CACACTGGGA CAAAGCTAAA AAAGAGTTAG GTAAAAATGA 5760  
 AGTGACATTT TCAATGAACA CAGAAGATAC ACCAGATGCA AAAATATCTG CTGAATATAT 5820  
 20 CAAATCGCAA GTTGAGAAAA ATTTACCAGG AGTTACTTTG AAAATTAAGC AATTACCGTT 5880  
 TAAACAAAGA GTATCACTAG AACTGAGTAA CAATTTTGAA GCATCACTTA GTGGTTGGTC 5940  
 25 TGCAGATTAC CCTGATCCTA TGGCTTATTT AGAAACAATG ACCACAGGTA GCGCACAAAA 6000  
 TAATACAGAC TGGGGTAATA AAGAATATGA TCAATTACTT AAAGTAGCAA GAACCAAATT 6060  
 GGCACCTCAA CCGAACGAAC GATATGAAAA CTTGAAAAAA GCAGAAGAAA GTTCCTAGG 6120  
 30 AGATGCACCG GTAGCACCAA TTTATCAAAA AGGTGTtGCA CATTTaACAA aTCCTCAAGT 6180  
 AAAAGGATTA ATTtACCATA AATTGGTCC AAATAACTCA CTTAAACATG TATATATTGA 6240  
 TAAATCGATA GATAAAGAAA CAGGTAAGAA GAAAAAATAA TATGCTTTGT AAATTAGGCT 6300  
 35 GGAGACATAT CTCCAGTCTT TTTGTGTTGG ATAAAAaCTT TGGGAATAAA AATTTAAAAT 6360  
 AAGTCGTTTT TTAAATTACT GAAATTGATT AAATGCATAA ATAAGTGAAT ATTCTAAAAA 6420  
 TAAACTTGTA ATAATTTTTT CTATGAGTAA ACTAAAAAGA AAAAATTAGA TTGAAAGTAG 6480  
 40 GAGGCATATG TATGGGGAAG CTAATTAAAT ATATTTCAAT ACTTCTTATT GTCGTTTTAG 6540  
 TGTTGAGTGC TTGCGGAAAA AGCAGTAATA AAGATGAAGG AGTAAAAGAT GCTACTAAAA 6600  
 45 CGGAAACCTC AAAACATAAA GGTGGTACCT TAAATGTAGC ATTAACAGCA CCGCCAAGTG 6660  
 GTGTTTATTC TTCGTTATTA AATAGTACAC ATGCAGATTC TGTAGTTGAG GGATATTTTA 6720  
 ACGAAAGCTT 6730

50 (2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6482 base pairs  
 (B) TYPE: nucleic acid  
 55 (C) STRANDEDNESS: double

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

5	AATTTTGTGC ATTATTAAAA ACCTCGCTTT TAAAAGATTG AAAAGTAAAT GAGTGAAATT	60
	AAAGATTATG CACATTAAAA TCACGCCACA ATTTAATTGT GAAAAATATC ACAAATATAT	120
	TATAACACTA AATTTCCCAA AATTCAAAAG TGTGTTTTAT TGCAGAAAAC TTATAACAyG	180
10	TGCACAAGTT ATAGTGAATT GCAAACGGAT TACTTTAGTC TTTTAAAAC ATGAAGTATA	240
	ATTTGTATAG CAATAAATAT AAAAATGGGA GGCTATGTTT AATGAGCAAT ATGAATCAAA	300
	CAATTATGGA TGCATTCAT TTCAGACATG CGACTAAGCA ATTCGATCCA CAAAAGAAAG	360
15	TTTCGAAAGA AGATTTTGAA ACAATATTAG AGTCAGGTAG ATTGTCTCCA AGTTCTCTTG	420
	GGTTAGAACC TTGGAAGTTT GTCGTGATTC AAGATCAAGC GTTACGTGAT GAATTAAAAG	480
	CGCACAGTTG GGGCGCAGCA AAACAATTAG ATACAGCGAG CCATTTTGTG CTAATTTTTG	540
20	CGCGTAAAAA TGTAACGTCA AGATCACCGT ATGTACAACA TATGTTAAGA GATATTAAAA	600
	AATATGAGGC ACAAACGATT CCAGCTGTTG AACAAAAATT CGATGCATTC CAAGCAGATT	660
25	TCCATATTTT TGATAATGAT CAAGCCTTGT ATGACTGGTC AAGTAAACAA ACGTATATTG	720
	CATTAGGCAA TATGATGACG ACAGCCGCAT TGTTAGGTAT TGATTCATGT CCGATGGAAG	780
	GTTTTAGTCT GGATACAGTG ACAGACATTT TAGCAAATAA AGGGATCTTA GATACTGAGC	840
30	AATTTGGTTT ATCAGTGATG GTCGCATTTG GCTACAGACA ACAAGAGCCA CCGAAAAATA	900
	AAACACGCCA AGCTTATGAA GATGTTATTG AATGGGTTGG ACCAAAAGAA TAAATAGAAT	960
	ACCGTATGTC TAAATATATA AAATTAAAAA GTTAGCAATA AAAAAGCCTG CGATTACATA	1020
35	AATGAATCGC AGGCTTTTGC GTGAAAAAAT TGTATTAATA AAGTATGGAT GATTATTTTT	1080
	CTGGGACAAG GTCAGTATTT GAATGAACTG TGATGTCAAA CCCTTCTGGT GCCGTAAATG	1140
40	TATGTGTTGA GGCCTCGGGT TGATAAATAT CAACATGTGT TAATCCATAA CTTTGTGAAT	1200
	TGTTTTGTCT TGCTTGATTG GATTGCCAAG TATTAGCAGC AATATGATGG TGATAATGAT	1260
	TCGTTGACAT AAATAGCGCA CGTGGAATAT CAGACACATG TTGGAATCCT AATTGTTCAA	1320
45	TGTAACATTG ATATGCTGCG TCTAAATCAT GTGTTTTTAA ATGTAAGTGT CCAATCATGC	1380
	CTTTTGCTGG CATTCTTGC CAACCTTCAT CAGTACGATG TGTTAATAAG GTTTGGCTAT	1440
	CAACTTCTAA AGTATCCATT TTAACCTTGC CATTTTGCCA TTCCCATGAA GATGAAGGTC	1500
50	TATCGCGATA GACTTCAATA CCATTACCTT CGGGGTCGTT GAAATATAAA GCTTCACTTA	1560
	CTAAATGATC ACCAGCGCCG ATGCCCATAT TTTTTTGTGC CACGAAATAT AAGAAGTTAG	1620
55		

	aAGTCTGACG GcCGTCTTCT AATAAATGTA ACGTTAGAGT ATGGcCACCA GTCCCAACAG	1740
	ATAATACGGT TGTATTATCG TCAGAACTTT TAACGGATAG TCCTAAAATG TTTTGTAAA	1800
5	ATGTTGTCAT TAAGTCTAAG TCTCTTACGT TCAGTACAAT GTTTGTCACT TGTGTTGCTG	1860
	TTTTATCGTG AAATGCCATT ATGCATCGCC TCTTTTCTA TTTTCTATA AGTTAGTATA	1920
10	AAAAGTATAC CAGAAAAGAA AATGAATTGA TAGCATAAAG TTTGAAATGC AAAATAACTA	1980
	GTCGTTTTGC AATTTTATcAT TGATGCGAAC AAAAAAGCGA TGGTACAGTT GCACCATCGC	2040
	AAAATTTATT TAACCAAGAT ATACATCTTG ATATGAATCT TCTTTTTCTA ACATATGTTT	2100
15	GGCAAATGAA CATGAGGCAA TAATTTTCAA ATTATTTTCT CGAGCGTGTT CAACAACTGc	2160
	TTTAAGTAGT TTTTGGCAA CACCTTGACC ACCAAGTTCA TCAGATACGC CTGTATGATC	2220
	AATGTTAATT TCATTATTAT CCACAAAACG GTATGTGATT TCAGCTAAAG CATTATTTTC	2280
20	ATCATCACCA ATATAGAATT TGTTCGCGC TTGTTTGATT TCAAGGTAC TCATACATAT	2340
	CAACTCCTAT CATGATTGAT TATAGTATTT CCCTATTCTA TTTTAACTTA AACGAAGTCA	2400
	AAGGTGCATG ACAGTCATGT GACGACATTG CCACATCTAT GTAGTCGTTT TTATTAAGCA	2460
25	CAGTTTGAAA TGAAGATGAA AACACGTATC TTGACATTAA ATCTATTCAG CTATATAATT	2520
	TATCTCGAAA TCGAAATAAA ATAAAAAGT TGGTGATCAT ATGGATCGAA CGAAACAATC	2580
30	TCTCAATGTT TTTGTGGGAA TGAATAGGGC GTTAGACACA TTAGAGCAA TTACAAAAGA	2640
	AGACGTAAAG CGATATGGCT TAAATATTAC TGAATTTGCA GTGCTCGAGT TGCTTTATAA	2700
	TAAAGGTCCG CAACCAATTC AACGTATTAG AGACCGCGTA TTAATTGCAA GTAGCAGCAT	2760
35	TTCATATGTT GTAAGTCAAT TAGAGGACAA AGGTTGGATT ACACGTGAAA AGGATAAAGA	2820
	TGATAAACGT GTATATATGG CTGTGTTAAC TGAAAAAGGT CAAAGTCAA TGGCAGATAT	2880
	TTTcCCTAAG CATGCTGAGA CATTAAACAA AGCGTTTGAT GTGTTAACAA AGGATGAATT	2940
40	AACAATCTTA CAACAAGCGT TTAAGAACT AAGTGCACAA TCTACAGAAG TGTAAGGCGT	3000
	GCACTAAAAA TTTACATTAA AGTATCTCGA TTTGAGATA AATGCACTAA AAATATAAAG	3060
	AGGGTATATA AAATGATAAA TAATCATGAA TTACTAGGTA TTCACCATGT TACTGCAATG	3120
45	ACAGATGATG CAGAACGTAA TTATAAATTT TTTACAGAAG TACTAGGCAT GCGTTTAGTT	3180
	AAAAAGACAG TCAATCAAGA TGATATTTAT ACGTATCATA CTTTTTTTGC AGATGATGTA	3240
50	GGTTCGGCAG GTACAGACAT GACGTTCTTT GATTTTCCAA ATATTACAAA AGGGCAGGCA	3300
	GGAACAAATT CCATTACAAG ACCGTCTTTT AGAGTGCCTA ACGATGACGC ATTAACATAT	3360
	TATGAACAGC GCTTTGATGA GTTTGGTGTT AAACACGAAG GTATTCAAGA ATTATTTGGT	3420
55		



	TTAAATGAAG GGGTAGCACC TGGTGTACCT TGGAAGAATG GACCGGTTCC AGTAGATAAA	3540
	GCGATTTATG GATTAGGCCC CATTGAAATT AAAGTAAGTT ATTTTGACGA CTTTAAAAAT	3600
5	ATTTTAGAGA CTGTTTACGG TATGACAACT ATTGCGCATG AAGATAATGT CGCATTACTT	3660
	GAAGTTGGCG AAGGAGGCAA TGGTGGCCAG GTAATCTTAA TAAAAGATGA TAAAGGGCCa	3720
10	GCaGCACGTC AAGGTTATGG tGAGGTACAT CATGTGTCAT TTCGTGTGAA AGATCATGAT	3780
	GCAATAGAAG CGTGGGCAAC GAAATATAAA GAGGTAGGTA TTAATAACTC AGGCATCGTT	3840
	AATCGTTTCT ATTTTGAAGC ATTATATGCA CGTGTGGGGC ATATTTTAAT AGAAATTTCA	3900
15	ACAGATGGAC CAGGATTTAT GGAAGATGAA CCTTATGAAA CATTAGGCGA AGGGTTATCC	3960
	TTACCACCAT TTTTAGAAAA TAAAAGAGAA TATATTGAAT CGGAAGTTAG ACCTTTTAAT	4020
	ACGAAGCGTC AACATGGTTA ATTGGAATGA GGAGGATTTG TGATGGAACA TATTTTLAGA	4080
20	GAAGGACAAA ATGGTGCGCC AACACTAATA TTATTGCATG GTACAGGTGG TGATGAGTTC	4140
	GATTTATTAC CGTTAGGCGA AgcATTGAAT GAAAATTATC ACTTGTTAAG TATTAGAGGA	4200
	CAAGTTTCAG AAAATGGGAT GAACCGTTAT TTCAAACGTC TTGGTGAAGG TGTTTATGAT	4260
25	GAAGAAGATT TGGCATTTCG TGGACAAGAA TTGTTGACGT TCATTAAAGA AGCTGCTGaA	4320
	CGTTATGATT TTGaTATTGA AAAAGCAGTA CTTGTTGGAT TTTCAAATGG ATCAAATATA	4380
30	GCGATTAAct TAATGTTGCG TTCAGAAGCA CCATTTAAAA AAGCATTGTT ATATGCACCG	4440
	TTATACCCAG TTGAAGTAAC GTCAACAAAG GATTTATCAG ATGTCAGTGT GTTGCTTTCT	4500
	ATGGGGAAAC ATGATCCAAT TGTGCCATTA GCTGCAAGTG AACAAGTCAT TAACTTGTTT	4560
35	AATACACGTG GGGCACAAAGT CGAAGAAGTT TGGGTGAAGG GCCATGAAAT TACAGAAACT	4620
	GGATTAACGG CTGGTCAACA AATACTTGGG AAATAACAGT TCTATTAAGA AGCGGACAGA	4680
	TGGA <sup>~</sup> AAAGAT TTTTACTTTT CATCTGCCCG CTTTTTTGAT TTTGAAGTGC TGTAATAAT	4740
40	TTTACAATAG TATAGATATT TTAATCGATA TGAGATTGTC CGGTAATACG CTTAATTAAA	4800
	CCTTTATAGA GTACAGGTAT GAGTAAGATG AAACCGAACA ATCCCATAAT AGGGAATACT	4860
	TTTCCAATTA ATGAAATGaA ACCGATAAAT GACTAATAT AAGTGATGAC AGCCATTGTA	4920
45	ATAATAATGA TGAAGTAACG TCTGCTGAAT GGAACGCTGA AACGTGACGC AAATGCATAC	4980
	ATTAATCCAA CAACAGTATT GTAGATGACA AGTATCATAA TGACAGACAT AATAATACCA	5040
50	ATTGACGGAG ACATTTGTGT CGCTAATTTT AATGTAGGTA GATCTACGTG TTTAATTTTA	5100
	TCGAATTGAG AAATTAAACC TAGATTAATC ATCATGAGTA AAAATGTAAT GATTAAACCG	5160
	CCAATCAAGC CCCCCTATAA CGTTGAGTCA CGATATTAA CTTTACTACC CATCACTGAT	5220
55		

CCAGGTGATA ATGATTTCTG CTTATGAATC TGAGCATCAT TATTAGCGGC AGTAAATCA 5340  
 AGATGACTTG TTGTGAAATA GTAGACCGCA ATCATAATGA CAATCGCAAT TAAAAATGGG 5400  
 5 GTAACACCGC CAAGCACAGC AATTAAACGA TCGAATTTTA GAAACAGTGT TGCTAAAATA 5460  
 AAGGCGACTA ATATGAGTGC GCTCAGCCAA TACGGTAACT TGAAACTTTG ATGAATGGTT 5520  
 GACGCACCAC CTGCAGTCAT AATAATAGCT AAAGACAACA TAAACATTGT TAAAATAATA 5580  
 10 TCAAAACCTC TTGCAATAGA GGGGTATAAG AAATAGTTAA TTGAATCAGA ATGATTTCTG 5640  
 GACTTTAGAT GATGACCTGT ATGCATGACA ACCATTCCAC CTAAAGTAAT CAATAGTCCT 5700  
 15 GTTACAATAA TGCCTGAAAT GCTATATGCG CCATGACTTG TGAAAACTG GAAAATTTCT 5760  
 TGACCAGTAG CAAAGCCGGC ACCAACGACA ACACCAACAA AGGCAAATGC CACAATAATG 5820  
 GACTCTTTTA AGATACGCAT GATTTAAAAA TGTCCCTTCG TAATTTTAAG TAATATAGAA 5880  
 20 AATGTAACAT ACATGTTAAT GAAAAATATA GTACTAATAT AGTATTTTGT TAAATTGGAG 5940  
 TAGAAGCGAG GGTGTCGGTC ATTTCAATTAA TTTATTAGTT GATTTTGCAT TTTTTTGCTG 6000  
 TAAAGTTGTT ATAATACAGT TAACAGGAAT TAGCATAGAT ACACCAATCC CCTCACTACT 6060  
 25 CGCAATAGTG AGGGGATTTT TTTCGGTGTA GCTAGGTCGC CTATTTATCA TCGTGTTTGC 6120  
 GTAgCaATGC GTAAACACAG TACCACTAAA TAAGTGCACG ATACATGCAT CAAATGTCGT 6180  
 CTTTAGTcTA AGTAACGATC ATGCATTAAC ATTTTCAAAA TATCTATTTG AGCTTGAAGA 6240  
 30 TCTTTACCAA TATTGGTATC ACGAATCTTC TTACGTTGTA ATTCTTTATC TACGACGCGC 6300  
 TTTATAGAAA GTTCATCGAT ACCTTCGGAA AGTATTTTTn CTTTAGCGTT AAATTGTTGG 6360  
 35 TGTGCAACGA GTTGCAATACC GAATGAATTA TACAATAGTG TATAGCCTGC AATGCCAGTh 6420  
 GTTGACTGAT AAGCTTTTGA AAAGCCACCA TCAATGACAA GCATCTTTCC ATCAGCCTTG 6480  
 AT - 6482

40 (2) INFORMATION FOR SEQ ID NO: 53:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16592 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

ATTTAAGGCG ATTGCTTGTG TATTTCTCTC TTTTGTAGGC AAACCTGCAC TCGTTCCAAA 60  
 AAATGTAAC TCCATATATG CCCCTCCTTT TCTTCAATTC ATTTTATCAT AAAATTGTA 120

55

	AATTTTCTA	ACTTTAACGT	AGACATAACT	ATATAAATTT	TGATAATTAC	GTTATACTTA	240
	TCATTAATAA	GTATCACATT	AAACATGATA	CATGAATCGA	TATTTTCATTT	AAGACACTGC	300
5	ATACAGTCGA	GCATATTGTA	TGACCTACTG	AATGGATTAT	CTTATAATAA	TAAATCATAT	360
	ATCTAATTAA	GAATTGAGGT	TTTAATCTTG	AGTACTAAAA	ACAAACACAT	CCCATGTTTA	420
10	ATCACAATCT	TTGGTGCACT	GCGTGACTION	AGCCATCGTA	AGTnGTTTCC	ATCAATATTC	480
	CATCTCTACC	AACAAGACAA	TTTAGATGAA	CATATTGCCA	TcATCgGTAT	TGGACGTCGT	540
	GACAT'kwnTA	ATGATGATTT	CCGTAATCAA	GTAAAATCAT	CAATTCAAAA	GCACGTAAAA	600
15	GATACAAACA	AAATTGACGC	GTTTATGGAA	CATGTCTTCT	ATCATAGACA	TGATGTTAGT	660
	AATGAAGAAA	GCTATCAAGA	ATTACTAGAT	TTTAGTAATG	AATTAGATAG	CCAATTTGAA	720
	TTAAAAGGTA	ATCGACTATT	CTATTTAGCA	ATGGCACCAC	AATTC'TTTG	CGTTATTCT	780
20	GATTATCTAA	AATCTTCTGG	TCTTACTGAT	ACAAAAGGAT	TTAAACGCCT	TGTTATCGAA	840
	AAACCATTCTG	GTAGTGATTT	AAAATCAGCC	GAAGCATTA	ACAATCAAAT	TCGTAAATCA	900
	TTTAAAGAAG	AAGAAATTTA	TCGTATTGAC	CACTATTTAG	GAAAAGACAT	GGTTCAAAAT	960
25	ATCGAGGTAT	TACGTTTTGC	GAATGCGATG	TTGAACCAT	TATGGAATAA	CAAATATATT	1020
	TCAAACATCC	AAGTTACATC	TTCTGAAATA	CTAGGTGTTG	AAGATCGTGG	TGGTTATTAT	1080
30	GAATCAAGTG	GCGCGCTAAA	AGATATGGTG	CAAAACCACA	TGTTACAAAT	GGTTGcATTA	1140
	TTAGCTATGG	AAGCACCTAT	TAGTTTAAAT	AGTGAAGATA	TCCGTGCTGA	GAAAGTAAAA	1200
	GTACTTAAAT	CACTGCGTCA	TTTCCAATCT	GAAGATGTTA	AAAAGAACTT	TGTTGCTGGT	1260
35	CAATATGGCG	AAGGCTATAT	CGATGGTAAA	CAAGTTAAAG	CATACCGTGA	TGAAGATCGC	1320
	GTTGCAGATG	ACTCTAACAC	ACCTACCTTT	GTTTCAGGTA	AATTAACAAT	TGATAACTTT	1380
	AGATGGGCTG	GTGTACCATT	CTATATTCGT	ACTGGTAAAC	GTATGAAATC	TAAAACAATT	1440
40	CAAGTTGTCG	TTGAATTTAA	AGAAGTACCA	ATGAACTTAT	ACTATGgAAA	CTGaTAAACT	1500
	GTTAGATTCA	AACCTATTAG	TAATCAATAT	CCAACCTAAT	GAAGGTGgTA	TCTTTtACAT	1560
45	CtAAATGcTA	AGaAAAATAC	ACAAGGTATC	gAAACAGrAC	CTGtCCmATT	GtCTTACTCm	1620
	ATGaGCGcTC	aAGaTAAAT	GaATACTGTA	GATGCATATG	AAAATCTATT	ATTTGATTGT	1680
	CTTAAAGGTG	ATGCCACTAA	CTTCACGCAC	TGGGAAGAAT	TAAaATCAAC	ATGGAAATTT	1740
50	GTTGATGCAA	TTCAAGATGA	ATGGAATATG	GTTGaTCCAG	AATTCCCTAA	CTATGAATCA	1800
	GGTACTAATG	GTCCATTAGA	AAGTGATTTA	CTACTTGCTC	GTGATGGTAA	CCATTGGTGG	1860
	GGACGATATT	CAATAATTGA	ATTAAAACGC	ACATGTTAAA	CAAAAATAAA	TGAGCGAATG	1920
55							

	TATATTATGA AATTATATTT TACAATGCCC AAAACTATTT TAATAATCAT TGAACAAATG	2040
	GGTGATAAT TTATAGAAAT AATGTAGAAT AAAAATAAAT GATTGAATTA ATTGGAGTGA	2100
5	AAGTTTTGGA CGTTATCAAG CAAATACAAC AGGCAATTGT TTATATTGAA GATCGTTTAT	2160
	TAGAGCCTTT CAATTTGCAA GAATTAAGTG ATTACGTTGG TCTTTCGCCA TACCATCTTG	2220
	ATCAATCATT TAAAATGATT GTCGGCTTAT CTCCAGAAGC TTATGCACGC GCGCGTAAAA	2280
10	TGACACTCGC TGCAAATGAT GTGATTAATG GTGCTACACG ACTTGTAGAT ATCGCTAAAA	2340
	AATATCACTA TGCAAATTCA AATGATTTTG CAAATGATTT TAGTGATTTT CACGGCGTAT	2400
15	CACCTATTCA AGCCTCTACT AAAAAGATG AATTACAAAT TCAAGAGCGA TTATATATCA	2460
	AATTATCAAC TACTGAGAGA GCACCTTATC CATAAGATT AGAAGAGACA GATGATATTT	2520
	CATTGGTTGG ATATGCACGA TTTATAGACA CTAAGTATTT GTCACATCCT TTTAATGTTT	2580
20	CGGATTTTTT AGAAGACTTG CTCATTGATG GTAAATTAAG AGAGTTACGA CGATATAATG	2640
	ACGTTAGTCC ATTTGAACTA TTGTATTATTA GTTGTCTCTT TGAAATGGT TTAGAAATAT	2700
	TTGTAGGTGT ACCAAGTGAA CGTTATCCTG CACACTTAGA AAGTCGATTT TTACCTGGCA	2760
25	AACATTGTGC GAAATTCAAT TTACAAGGTG AAATTGATTA TGCAACTAAT GAAGCTTGGT	2820
	ACTATATTGA ATCAAGTTTG CAGTTAACAT TGCCATATGA ACGAAATGAT TTATATGTTG	2880
	AAGTGATCCC TCTCGATATT TCATTTAATG ACCCATTCAC TAAAATTCAG CTTTGGATTC	2940
30	CTGTAAACA GAGTCCTTAT GACGAAGATT AAATAATAAA AAACAAAGAA GCCCCCTAAT	3000
	ATATCTATAG GTCTACAAAT GGCCTTAGAT TCTATTAGGG GGCATATTAA TATGTTAATT	3060
35	TAGTTCGATA ACACATGCTT CATATGGACG TAACTGTTTT AAATTAACCTT TGGCATCATA	3120
	ATTAAATAGC TTTACTTCTC CATGGCTTAA ATCAAATGGT ACAGTTAATT CTGCTTCGTG	3180
	GTTAGTAAGA TTACCTACAA TAAGAACTTG CTTTTCAITTT AATGTTCTCG TGTACGCAAA	3240
40	AACTTGTGAA TTTTCAGCAT CTAATAATC AAATTGACCA TATACGTATA CATCATTAGA	3300
	CTTTCTTAAT TGAATTAAAT CTTTATAAAA TTGTAATACT GAATGCTCAT CTTCTAATTG	3360
	TTGTGCAACA TTGATAGTTT TATAATTCCG ATTCACTGGG AACCACGGTT CACCATTGTG	3420
45	AAATCCTCCA TTTAACGTAT CATCCCATTG CATGGTGTG CGAGAATTAT CTCGGTTCTC	3480
	ATCTTTATAT TTCGCAAGTA AAGCGTCTAC ATCTCCACCT TGAGCTTTCA CTATTTGATA	3540
	GTCATTTTTA ACAGCAACAT CGTTAAACGT TTCAATACTT TCAAATGGAT AATTCGTCAT	3600
50	ACCAATTTCT TGACCTTGAT AAATGAATGG CGTACCTTGT TGCAAGAAAT AAACAGCTGC	3660
	ATGACTTGTT GCTGATTCAT ACCAATACTT GTCATCGTCA CCCCACGTCG ATACACGTCG	3720
55		

	CCATCTATTT AATACAGATT TATACGAATT TACATCAAAG TGAGAATCAC CACTATTCCA	3840
	CAGTCCCAAA TGTTCAAATT GGAATATCAT ATTAAATTTA CCATTTTCTT CCCCAGCCCA	3900
5	GTCATCAGCA TCATCAGGGC TTACACCATT CGCTTCACCA ACAGTCATAA TGTCATACTT	3960
	ACTTAATGAG CGATCTTTCA TCTCTTGTA CCAAGTTTGT ATACCTGGCT GATTCATATC	4020
	TACATCAAAT GCTGGGGCAT ATGTTTTACC CTCAGGTACA GGTAAGTCAC CCGCTTCAAA	4080
10	CGTCTTCTTA ATATGCGTAA TTGCATCTAC TCTAAATCCA TCAATGCCTT TATCAAACCA	4140
	CCAGTTCATC ATTTCAAATA CAGCATCTCT AACTTCCGGA TTACCCCAAT TCAAATCAGG	4200
15	TTGTTTTTTA CTGAATAAAT GGAAATAATA TTGCTCAGTA TTAGCATCAT ATTCCCATGT	4260
	AGATCCATTA AATATACTTT CCCAGTTGTT AGGTTGAGAG CCATCTGGCT TTGGATCTTG	4320
	CCAAATGTAC CAATCACGTT TGGGATTGTC TTTACTAGAT TTGGATTCTA TAAACCAAGG	4380
20	ATGTTTCATCA GATGTATGAT TTACAACTAA ATCTAAAATA AGCTTCATGC CTCTATCATG	4440
	AACACCTTTT AATAAACGAT CAAAGTCTTC CATCGTTCCA AATTCATCCA TAATCTCTTG	4500
	GTAGTCACTA ATATCATAAC CATTGTCATC ATTAGGTGAT TTAAACATTG GACTGAGCCA	4560
25	AATGACATCG ATACCGAAAT CTTTTAAGTA GTCCAATTTA TCAATCATTG CAGGTAAATC	4620
	CCCAATACCA TCGTGATTAC TATCATTAAA ACTTCTTGGA TATACTTGAT ATGCTACTGC	4680
	TTCTTTCCAC CATTGCTTAT TCATTTTAAA ACTCCTTGC TATCGCTGTG TTGATTTTCT	4740
30	TATTTTAAAT TCTGTATCTA TAATGACGAG TTCAATAACA TCCTGTGCTT TGTTTTCAA	4800
	TATATTTAAA ATTGCTGCAC CAGCCTGTTG ACCTAACATT CGAGGCTTGA TGTCAATACA	4860
35	GGTTTGTGGT GGTGACGCAA TTTCGGTTAA ATAAGAATCA TTGAACGTTG CTGTCATTAC	4920
	ATCTTTCCGA ATTTCAATAT TAAGTTCATA TAGGACACTT AAAATCGCTA AATGTAACAT	4980
	AGCATCTAAC GAAATGATTG CCTGTTTAAAT ATTTGGGTCC TTCAAACGCG TATGTAGATT	5040
40	TTGCATGTAA TTAAAAATAA CTCTCTTTC ATTACTAGTC TCAATAATTT GATAATTAAT	5100
	TTTATTTTGA GAAGCTATCG TTTCAAATCC TTGAATTCTA TCTTTTGAAA CTTCAAAT	5160
	TCCTTTTCT GTAATAAATA TTAATTCATC TACACCTTGT TCAATAACAT GTCGTGTCAA	5220
45	ATTTTCAGAA GCTAATATAT TATCATTATC TATATGTGTA AATTGATGAT CTATATCCGA	5280
	TGTAGGCTTA CCAATCACAA TAAATGGCAT GCTTTCATCA ATTAACATTT GTTTAATCGG	5340
	ATCATTTTCT TTTGAATAGA GCAGTATAAA CGCATCAACC ATTCGTTGTT TAATCATTTT	5400
50	ATAAACTTCA TCCATTAAAT CATTCAATAT ATTTGAGACT GTCGTTTGTG TACCATAGCC	5460
	ATGCTGGTTA CACGTTTCAG AAATTCCTAG CAATACATTG ATGTAGAATG GATTCAGTCG	5520
55		

	AGTTCTAGCA GCGGTATTAG GAAAATAATT CAATCTTCC ATAACCTTCT TCACTTTTGA	5640
	AATTGTGCGT TCGCTAATAC GTTGATTTCC TTTTATAACT CTTGAAACTG TCGAAGGAGA	5700
5	AACACCGGCT TTAGTGCAA CATCTTTAAT CGTAACCATT TAATCACCTC CTGTTAATTT	5760
	CTGCATCGGA AAACGCTTCC AACCCTGTGA TAATACCAGT TTAGTCACAC TTTCTAAAAA	5820
10	AGTCAAAAGA TTTGTGCAA CGATTGCATA AAACGATAAA AATAAAACCT TCATACTGAA	5880
	ATTCAATCCG AAAATCAATA TAAAGGTTTG TATAAATATT AAAATCGATT GTTTAGTCAC	5940
	TAACTGCAA ATAGTTACCT TGGCCATCTT GAAAATTAAA TACACGTTGA CCATTCATTT	6000
15	CTACTATATC ATGCCCAGTT AAACCTAAAT CATTTAATTT TGAGTATAAT GCATCAAAGT	6060
	TTTTCTCTTT AAACATTAAA GATGGTGTTC CTAGGTTTAC TTCCGGGCTA TGCTTTTCAA	6120
	TAAATTCTTT TGCCATAATC GTCAATGACG TTTGAGCATC TTTGGTAGGT GATACTTCAA	6180
20	CTGCAACATA GTCCTCAGCT AACGGTGTTC CACTTACAAC AACAAATTCT AAAGTTTCTG	6240
	TCCAAAATGC TTTGCTTTT TCGACATCAT CAACATATAA CATAACTTGA TTTAACTTTT	6300
	CCATAAAATA GTACCTCTAT TTCTCTATAG TACATGCTAT CATAACACAG TAAATATTTT	6360
25	ATTACTTCAC AAAATGCTTA AAAATATGGC GGGATGCTTT TAAGGTCAAG GATAATACTT	6420
	GTGTAATTTT TTATAGGTTG TAGCTACTCT ATCACACTCT CTTTTATATT TATCAAAAGA	6480
30	TATAAAAAAG GATAGTATCT TTCAACTATC CTTTAATCAA TATTATTCTT CAATCCATTG	6540
	TGTATGGAAT ACGCCTTCTT TATCTTTTCT TTCGTACGTA TGAGCACCGA AGTAGTCACG	6600
	TTGTGCTTGA ATTAAGTTTG CAGGTAAATC AGCAGCACGG TAACTATCAT AGTAATTAAT	6660
35	ACTTGATGAG AAACCAGGTG TTGGTACACC ATTTTGAACA CCAGTTGCGA CAACATCACG	6720
	TAACGCATCT TGATATTCAG TAACGATGTT TTAAAGTAA GGATCTAGCA ATAAGTTTGT	6780
	TAATCCTGGA TTATTATCGT AAGCATCTTT GATCTTTTGT AAGAATTGTG CACGGATAAT	6840
40	GCAACCTTCT CTCCAAATCA TAGCTAAATC ACCAAGTTT AAATTCCATT CATTATCTTC	6900
	ACTTGCTTTA CGCATTTGcG CGAAACCTTG TGCATAAGAA CAAATTTTAC TCATATATAA	6960
	TGCTTTACGA ATTTTTTCTA AAAAGTCTTT CTGTGCACCA TCAAATGATG CTTTGGACC	7020
45	ATTTAATTCT TTAGAAGCAT TTACGCGCTC TTCTTTGaTT GAAGAGATAA AACGTGCAAA	7080
	TACAGATTCA GTAATGATTG TTAATGGAAT ACCTAATTCT AATGCGTTAA TTGAAGTCCA	7140
50	TTTTCTGTGTA CCTTTTGTaC CTGCAGTATC AAGAATTTT TCAACTAATG CTTCTTTATT	7200
	TTCATCTAAT TTCATGAAAA TATCACCAGT GATTTCAATT AAATAACTTT CTAATTCACC	7260
55	AGCATTCAG TCTTTGAACG TTTGAGCAAT GTCTTCATGA GACATGCCTA ATAATTCTTT	7320

EP 0 786 519 A2

	CATTTTCACA TAGTGTCAG CACCAATTAGG TCCAATATAA GTAACACATG AAGCACCGTC	7440
	TTTTGCCTTT GCAGCAATTG CATCAAGAAT ATCTGCAACT TTGTTATAAG CTTCTTCTTG	7500
5	TCCACCCGGC ATTAATGACG GACCAGTTAA CGCTCCAATT TCACCACCAG AAACGCCCAT	7560
	ACCAATAAAG TTGATTGCAC TTTGTGyWAA TGCTTTATTA CGTCTGATAG TATCTTGATA	7620
	GTTTGTATTA CCACCATCAA TTAAAATATC TCCATCATCT AATAAAGGTA ACAAACATC	7680
10	AATCGTTGCG TCCGTAGCTT TACCTGCTTG AACCATTAAT AAAATTTTAC GTGGTTTTTC	7740
	TAAAGAATTA ACAAATCTT CCAATGAATA CGTTGGATGA ATATTTTTCC CTTTTGATTC	7800
15	TTCAACCATT AAATCAGTTT TTCACTTGA GCGGTAAAT ACAGATACAC TATATCCGCG	7860
	TGATTCAATA TTCCAAGCTA GGTTTTTACC CATAACGGCT AAACCAATAA CTCCAATTTG	7920
	TTGTGTCATA TTACTIONCT CACTTGTTGA TTTTTCATTA GTATTGTATC ACAAATAGA	7980
20	CATACACTAC ACTAAATCAT TTCGAATGTC GCGCAACTAT TTTGATTATT TCTAACACTT	8040
	GACTTGCAAG CAAGTTCAAT GATTTAATCG GCATTCTCTC ATTTGTTGTA TGGATTTTTT	8100
	CATAACCCAC TCCTAAAATG ACTGAAGGAA TACCAAATGT ATTAATAATA CTGCCGTCTG	8160
25	AACCGCCACC AGAAATAATT GTATTTGCAG ATAATCCTAA ATTACGAGCA CTTTCTTGTC	8220
	CAATTTTAAC AACCGCTTCA TTATCATTAA TTTTAAATCC TGGATAACTT TGCTCCACTG	8280
	TAACACTGTC TTTCCACCT AATTCTGATG CAGTAGTTTC AAACACATCA GTCATATGTT	8340
30	TGACTTGTGT TTTTATTCTT TCTGGATCGT GAGAACGTGC CTCTGCTTCT AAAATGACTT	8400
	CATCTGCAAC AATATTGTA GCTGAACCGC CATGAAACTT ACCAATATTG GCAGTAGTTA	8460
35	TTTCATCAAC TTGTCCTAAT TTCATTGAC TAATTGCTTT CGCCGCAATA TTAATAGCAC	8520
	TAACACCCCTC TTTTGGCGTA CTGTCATGAG CCGTTTTGCC AAAAATTTTA GCTGAAATTA	8580
	ACATTTGCGT CGGTGCACCT ACAACCGTAG TACCGACATC AGCACTTGCA TCAATAGCAT	8640
40	AACCAAAGTC CGCGTCCAAC AACTCTGAAT TTAATTCTTT AGCACCAATT AAACCTGATT	8700
	CTTCTCCAAC AGTAATCACA AATTGAATTT GTCCATGTGG GATTGTGTTG TCCTTTATCA	8760
	CTTGCAAAAC TTCAAGCATC GCTGATAATC CTGCTTTATC ATCTGCACCT AGAATAGTCG	8820
45	TACCATCAGA GTATATGTAG CCGTCATCTT TTACAATTGG CTTTACATTA ATTGCGGGTA	8880
	CAACAGTATC CATATGGCTC GTCAAATATA ATTTAGGTAC TTCGCCTTCT TCGATAGTAC	8940
	TATTCATTGT ACACACTAGA TTATTGGCAC CTAATTTAGG ATGTTTAGCC GCTTCATCTT	9000
50	CTTTAACATC TAACCCTAAT GCTATGAATT TTTCTTTTAA AATAGGTTGG ATTGTTGATT	9060
	CATTCCCTGT CTCAGAATCG ATTTGTACAA GTTCAAAAAA CGTATTAAGT AATCTTTGCT	9120
55		

	GATGAAATAA AATGTTACAG TAATTGACGT TACACAGATT TATCAGGTTT GTAAATTGTG	9240
	TCATATTATT TTCAATTTAT TATATATAAT TATTGTAAC TAACTAAGC TTTGTCAAAA	9300
5	ATATATTGAT TGATTTTTCA AAGATATCGT ATAATGAGGA AAATGACATA AGCAAACTTA	9360
	CTCATGTTTT TTATTATATT CCTTTATGAT GATTGCTAGT TATATCGTCT CAAGTTAAAA	9420
	GTTTTATATC TTATGTCGTA ATTATTAATA CAAAGGTTAT TCATTTGGAG GCACACAAAA	9480
10	TGCAAAATAA AGTTTTAAGA ATTATCATT TCGTTATGCT TGTATCAGTT GTATTAGCAT	9540
	TGTTATTAAC GAGTATCATT CCAATTTTAT AAACATATC TCAACTACCT ATACAAAATC	9600
15	ATACAATTAA AAATCCATCC ATTATAAAG CATGTATTAA TAAGTTATCG TATTGCAACG	9660
	ATTACTTTCA AACATGGGTC ATACGGATGG ATTATTTTTT AAGCTACTTC ACTATGCATT	9720
	TTCAATGAAC CAAATGCGA TTTGATTTGT AAATATTCTT CTAATTCATT TAATATTTGA	9780
20	ATAATACTTG CTCTCGAGTT AAGCGCTTTG TGTGTTGTTG GCAATGGCAG TTCATCCAAT	9840
	TTCAAACGCG TCTCATACAA ATTGTGTAAC CGCATTGCTG TATAGTCATT ACTATTCACA	9900
	TTTAGACCAA TTTCTTTCAG CAGTGACGCA ACATCATTTA AAAGCGGATC TTTATGACAG	9960
25	ATACTTTCGA TGAGCGGTTT CATTCCTCATT AACAATTCCA CTTGCTCTTC TCGCATATCA	10020
	AAATAATGAT AGTATGAATT TTCGTTTCTA ACAAATGAT TTTTAACATC TCGGAACGCG	10080
	ATAGACTTCG CCTTTTAAAT ATTTAAAAGT AACACTTCAA ATTCAATCGC AATGGTATCT	10140
30	TCATATTTTT CACAAATATA ACTATATTTA CTAAAAATAT CAGCAATTTG TTGCTCAATT	10200
	TTACATTTGT ATTCTGCTAG TTGTTGTCT AACTTGGCA TCATTAAATT CATGTAAAT	10260
	GCAATGCTTA GTCCAATTAA CAGTAATAAT GTTTCATTAA CAATTAAATG TGCATCAATT	10320
35	GATTTTGCAT TAAAAACATG AAGTAATATA ACGCAACTCG TAATGACACC TTCTTGTAAT	10380
	TTTATACGA CAGTTAATGG TATAAATAAC AATACGATAA TACCGAGTAC AATTGGACTC	10440
40	TGACCTAATA AACTAAATAT TGCTGAACCT AAAAACAATA CTAAAAACA TGATACTAAT	10500
	CTTGAAATAA TCGCTTGTA CGAATGTAAT TTTGTATGTT TAATACATAA TACGACTAAT	10560
	ATGGCGCTTG AAGCATAATT ATCTAAACCT AACAGCTTAC TAATAATTAC ACCTAAAGTC	10620
45	ATACCCACTG CTGTTTTTAT TGTCTAAAT CCAATCTTGT AAGGATTTAA CTTTAACATG	10680
	GGTTAGCGCC TCTTATCTTT CTTCAATAA TTTATTGAAT AATGTTTGTA ATTGATTAAT	10740
	TACGTTTCATC ACATCATGAC CTTGATTTG ATGTCTTTCA ATCATTTCTG TAATCTTTCC	10800
50	ATCTTTTACT AATGCAAATG ACGGACTTGA AGGCGCATAA CCTTCGAAGT ATTCACGCGC	10860
	TCTTTGTGTC GCTTCTTTAT CTTGTCCAGC AAATACTGTC ACTAGACGAT CAGGTAATAC	10920
55		



	AGAATTGATC ATAAC TAGTG TTGTACCATC TTGTTAAGA ACTTTGTCAA CATCTTCTGC	11040
5	AGTAGTTAAT TGCTCATATC CCGCAGATT C AATTTCAATC CTTGCTTGTT CTACAACACC	11100
	GTTTCATGTAT AAATCGAAAT TCATGnCCAT AAGTTCAATC ACCTATCCCT TTATATTTAA	11160
	ACTAcCCTCA TTCTACTAAT TAATAACATA TTGTTC AATAA CTG AATCACACCT	11220
10	ATATTTAGAC ACAATTTTAA CAATATACCA AACATTATTG TGCTTAAAT CATGGTAACT	11280
	AATTTGTTCA CATGTTTTCA TTAATATGTT TCAAGTATGA TGTCTTATTT TGACTTTACT	11340
	GCAAAAATGC ATTCAACCAT GTTGATTATT GTTCTTTATC TTTTTTGAAT ATATTGCACA	11400
15	TATTTTAGTG CCAAAAAATA ATACATCCAT CGACAAGAAC AAGATAAAAC AAGTTGTCGA	11460
	TAGATGCATC TATGTTATCA CTAATATATA TTTGTATTTT CTAAAGTATA CTGTTTCGATA	11520
	CGCTGTTTAA TATGATTCAT AaATTTACCT GTTTGTAAAC CATCTAAAT ACGATGATCA	11580
20	ATTGAAATAC ATAAATTAAC CATGTTACGA ATTGCAATCA TATCATTAA TACTACTGGC	11640
	TTTTTAACGA TTGATTCTAC TTGTAAATC GCTGCTTG TGATGATTTAT AATACCCATT	11700
	GATGATACTG AACCAAATGT ACCAGTATTA TTTACCGTAA ATGTACCGCC CTGCATATCT	11760
25	TCAGCTGTCA ATTGCTTATT ACGCGCTTTC GTTGCTAAAG TATTAATTTT TCTAGCTATA	11820
	CCTTTGATTG ACTTTTCGTC TGCATGCTTA ATCACAGGTA CGTATAATTT ATTTTCATCA	11880
	GCAACAGCAA TTGAAATATT AATGTCTTTA TGTAAGACAA TTTCATTTCC TTGCCAGCTA	11940
30	CTATTTAATA AAGGATATGC TTTTAAAGCA TCTGCTACAG CTTTACAAA GAAAGCAAAG	12000
	AACGTTAGAT TATATCCTTC TTTATTTTTA AAGCTGTTTT TATAATGATT TCTCGTATTC	12060
35	ACAAGATTG TAGCATCTAC TTCAATCATC ATCCATGCAT GTGGAATCTC TGTTACACTA	12120
	TTAACCATAT TTTGCGCAAT TGCTTTACGC ACACCATTTA CTGGTATTGT GCTGTTTTCA	12180
	CTATFGTCTT CAGATGATTG GTTACTTGAT GTATCTACTG ATGTTGATTT TGTTTGAAC	12240
40	TGTTTGT CAG ATTGAGCTGT GGTACCACCA TTTTCAATAA CTGACATTAT ATCCTTCTTA	12300
	GTTACACGAC CTTCAAATCC ACTACCTACA ACTTGTGATA AATCAATGTC ATGCTCTGAA	12360
	GCGAGTTTAA ATACAACAGG TGAAAAGCGA CCATTATTAC GTGGTTGATT TTGTTTAGCA	12420
45	GTAGATGTCT GTTCCACTGT TGCACTAGCT TTTT TAGTAG ATTTCTGAGT ATGCTCATCC	12480
	ACTTTTGCTT GTATCTCTTC AGTTGTTTCA TTTGTCTTTT CATCAGCAGT TTCAATTTTA	12540
	CAGATAATTG TATCAATAGC TACTGTCTGC CCCGCTTCAA CTAAAATTTT TGTAATTGTT	12600
50	CCTGATATCG TGGAAGGGAC TTCAGCTGTC ACTTTATCTG TAATAACTTC ACATAATGGT	12660
	TCATATTCAT CAATATGATC ACCAACAGAA ACTAACCATT GTTCAATGGT GCCTTCATGA	12720
55		

	AATTCACGCA TTTTATTTAA GATTTTTTCT GGATTCATCA TAATTTTCATT TTCTAATACA	12840
	GGAGAAAATG GCATAGATGG TACAtCTGGA GCAGCTAAAC GCATGATTGG TGCATCTAAA	12900
5	TCGAACAAGC AATGCTCTGC AATAATCGCT GACACTTCTG ACATAATACT ACCTTCTAAA	12960
	TTATCTTCAG TTACAAGTAA AACTTTACCT GTATGTTTAG CACGATCAAT AATTGTTTCT	13020
10	TTATCTAATG GATAAACAGT TCGTAAATCA ACGACTTCAA CATTGATACC GTCTGCAGCT	13080
	AAAATATCCG CTGCTTGTA ACAATAATTG ACCATTAATC CATAACAAAA TACTGTAA	13140
	TCTTCACCTT CACGTTTCAC ATCTGCTTTT CCTAAAGGTA CAGTGTAAATA TTCTTCTGGC	13200
15	ACTTCTTCCT TTAAGAAACG ATAAGCTTTT TTATGCTCAA AGTACAATAC TGGATCATT	13260
	GATTTCGATAG ATGATAATAA AAGCCCTTTA GCATCATACG GTGTGGAAGG AATAACAATT	13320
	GTTAAACCTG GCGATGAAGC AAATATACTT TCAATACTTT GTGAATGATA TAGTCCTCCG	13380
20	TGAACACCGc CACCAAATGG TGCACGAATC GTTAATGGGC ATTGCCAATC ATTATTTGAA	13440
	CGATAACGCA TTTTCGCAGC TTCACTAATA ATTTGATTG TCGCAGGTAA AATAAAATCT	13500
	GCAAAATGAA TTTCTGCAAT TGGTCTTTTA CCTACCATAG CTGCACCAAT GGCAGTTCCA	13560
25	ACAATATTTG ACTCAGCTAA TGGCGTATCG ATAActCTGT CTTCAACATA TTTTGTGTC	13620
	AGTCCTTGAG TAGTACCAA TACGCCACCT TTTCTACCAA CATCTTCACC AAGAATAAAC	13680
30	ACATCTTAT TTTGTTGTAA TGCTAAGTCT TGTGCCTGcG TATCGCCTCT AAATAAGATA	13740
	ATTTAGCCAT TAGTTAAGAC TCCCTTCTTC GTACACAAAT GCATAGGCTT CTTGACACT	13800
	TGGATATGGC GCGTCTTCAG CAGCCTTTGT CGCTTTATG ATGATGTCTT TnAtgTCCGC	13860
35	TTCTATTTCT GCCAACCAAG CATCATCGAT AATGCCAGCT GAAAGCAACT CTTTTTTGAA	13920
	CTTTTCATTG CAGTCTGCTT TTTTAAGcGT TTCACGCTCT TCTTTCGTAC GATATTGGTC	13980
	GTCArCATCT GATGAATGAG CTGTCATACG ACTTGTTACT GCTTCAATCA AAGTTGAACC	14040
40	TTGACCAGAA ATAGCTCGAT CTCTTGCTTC TTTCATCGCT TTATACATTG CTAATGGATC	14100
	ATTACCATCT ACTTGTTTAC CATGTATACC GTAACCAAGT GCTCTATCCG ATAATTTTTC	14160
	AGCTGCGTAT TGTAATGAAT CAGGTACTGA AATTGCATAT TTATTATTTA TAATGACACA	14220
45	TACAAAAGGA AGTTTGTGTA CACCCGCGAA GTTTAAACCT TCATGGAAGT CACCTTGTT	14280
	TGAGCTACCT TCACCAACAG TTGCTGTTGC AATTTTCTTC TTACCATCCA TTTTAAAGC	14340
	TAAAGCAGCA CCAACAGCAT GGGGTATTTG AGTTGCTACC GGTGAACTTT GAGACAAAAT	14400
50	ATTCTTAGCT CTACTIONTAA AGTGTGATGG CATTTGTTTT CCACCAGAGT TAACATCGTC	14460
	TTTCTTTCCA AACGCTGATA AAAACGTATC ATACGCTGAG ATACCCATAT AAGTAACGAA	14520
55		

EP 0 786 519 A2

	AATCTGAGTT GCTTCTTGTC CTTGACCACT TACAACAAAT GGAATTTTAC CTGCACGGTT	14640
	CAATAACCAC AGTCTTTCAT CTATTTTCT ACCTAAATCC ATCCATTTAT ATATTACTTT	14700
5	TAGGTCTTCT TCGCTAAGGC CTAATGATTT ATAATCAATC ATGTTAAATC CTCCTATTTA	14760
	TACGTGAATA GCTCTACTTT CTGCTTTCAA TCCTAATTCC ATCAACACTT CAGAGATGGA	14820
10	AGGATGTGCG TGTGTTGTTA GTCCTAATTC TAATGCCGAG CCATTCATGA ACTGTAACAG	14880
	TGATGCCTCA TTAATCAATT CTGTTACATG TGGACCAATC ATATTAATAC CCACAATTTT	14940
	TTCAGTTGAT TGATCAATCA CCATTTTCGT ATACCCTTGG TTTGTGTCAT GGCTATCAAT	15000
15	CACTGCTTTA CCAATTGCTT TAAATGGTAC TTTAAACTT TTAACCTTCA TTCCCTCTGC	15060
	CTTTGCTTGT TCAATGTTTA AACCGATAGA AGCAATTTCA GGTGTGAAT AAATACACTT	15120
	AGGCATCATG TTATAGTTTA CTGGGATTGG GTTCCCTCA AACATATGAT CAACAGCCAC	15180
20	AACACCTTCT TTTGATCCAA CATGTGCCAA TTGTAATTTT CCTATACAAT CACCAGCTGC	15240
	ATAAATATGT TTATCTTCAG TTTGTTGAAA TTCGTTTCGT AAAATATGTC CTGATGTTGa	15300
	AAGtTTTATT TTAGTGTTGT TTAAACCAAT ATCTGATGTG TTAGGTTTTT TACCAATCGA	15360
25	TAGCAACACT TTATCTACTT TAATTATGTC TGAGGAAATT TCAAACGTAA CACCATCTTC	15420
	GTTAACATTT ATATCATTTT CAGAAAGTTT TATTCCCTCA TAGAATTTAA CACCACGTGC	15480
30	TGACAATGAT TTTTTTAATA GTTGTGAAGC TTGTTTACTT TCAGTTGGTA AAATTCTTTC	15540
	ACCTGCTTCT ATAAGTGTTA CGTCAACACC TAAATCTATC ATCAATGATG CAAATTCCAT	15600
	TCCGATAACA CCACCACCAA TAATACCAAT ACTTGATGGT AACGTCTTTA ATGATAATAT	15660
35	ATCATCGCTA GATAAAATTT TATCATGATC AAATGATAAG AATGGCAACT CTGCAGGCGA	15720
	AGAACCAGTT GCAATTAATA CAAATTGGTT GGGTAATAAG TCTGATTCAC CATCTTCATA	15780
	TTGACAGAA ATTGTGCCAC TTTGAGGTGA AAATATAGAT GTACCTAGAA TACGTCCCGT	15840
40	GCCATTATAA ATGTCAATGT GATTGTGTTG CATTAAATGC TTTACACCTT GATACATTTG	15900
	ATTAATAATG TCTTCTTTTC GTGCCAACAT ATTTTCAAAA TTAACATTAG CATCTTTGAC	15960
	ATCAACGCCA AACATTGCTG CCTGTTTTAC TGTTTGAAAT ACTTCAGCAG ATTTAAGCAG	16020
45	CGATTTAGTA GGAATACAAC CTTTATGGAG ACAAGTACCT CCTAATAGTT GTCGTTCTAC	16080
	TATTGCCACT TTTTACCTA ATTGAGACGC ACGTATCGCA GCAACATATC CTGCAGTACC	16140
	TCCACCGAGA ACGACTAAAT CATATTGTTT CTCTGACATG TTCTTACTCC TAACTAATGA	16200
50	TATATATCCA TTGAAAATTT ATTAATACAT AGTTTTCATG TCCATTAAAT ACCTATTTTA	16260
	CATGATTGTC TATTTAGTTT GAATGCACAT AAATAAATCC ATAAATGAGT ATTCAACACA	16320
55		

TAAATCAGTA ACACTTGCAC CTGAAATCAT TCGTGCAATT TCATCTACTT TATCATCGCT 16440  
 AATTA ACTCT TGA ACTTGTG TTGTTGTACG ATCATCTTTT GATGATTTCG AAATTAATAA 16500  
 5 ATGATGGTCG CTCATCGATG CAACTTGTGG TAAGTGAGAG ATACAAATAA CTTGTATATA 16560  
 TTCTGCTaTA TCTCGCATTT TCTCTGCCAT TT 16592

10 (2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13794 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

20 CCAATACAAC GTAAAAAGAT TGCTTGTGTT ATTAATGAGT TAGATAAAAT AATTAAAGGA 60  
 TTTAATAAGG AAAGAGACTA CATAAAATAT CAATGGGCTC CAAAATATAG CAAAGAnTTT 120  
 TTTATACTTT TTATGAACAT TATGTACTCA AAAGATTTTT TAAATATCG ATTTAATTTA 180  
 25 ACATTTCTTG ATTTATCTAT CTTATATGTA ATATCATCTC GAAAAAATGA GATACTAAAT 240  
 TTAAAAGATT TGTTTGAAAG TATTAGATT ATGTATCCTC AAATTGTTAG GTCAGTTAAT 300  
 AGATTAAATA ATAAAGGTAT GCTAATCAAA GAACGATCCC TTGCAGATGA AAGGATTGTG 360  
 30 TTAATCAAAA TAAATAAAAT ACAATATAAC ACTATTAAAA GCATATTCAC AGATACTTCC 420  
 AAGATTCTCA AACCAAGAAA ATTTTCTTT TAAATTTAAA CAGATTTACC TCTTGATAAA 480  
 35 ATAAATAAGC AATCATACTA CTTCTCAATT TAGTATAAAT AAAAATACAT AATTAACTTT 540  
 CTTTGTGTTT TATATTATTT CAATACCCTA CTATATATCA CAACACATAA ATTAAGCATG 600  
 ACACTCATTC AATTAGTTC ACCATTTTCGT GTTCCAATT TACTGAGTAT CATGCTTTTA 660  
 40 ATGTTATAAA CCTAATGCTT TAATAAATCG TGTTAATTCT TCTCGCATAC TGTCATCTTT 720  
 CAATGCATAT TCTATGGTAG TTTTAACGAA GCCTAATTTT TCTCCAACGT CATAACGTTT 780  
 GCCTTCGAAG TCATATGCAT ACACTTGGTT ATCATTATTC ATACGTTCAA TCGCATCTGT 840  
 45 TAACTGAATT TCGTTACCTG CGCCTTCTTT TTGCGTTTTT AAATAATCGA AAATTTCAGG 900  
 CGTTAATACA TAACGTCCCA TAATAGCTAG GTTTGATGGT GCCGTACCTT GTGCTGGCTT 960  
 TTCAACAAAC TTTTTCACCT CATACTGACG TCCGTTTTTA GTTAATGGGT CAATAATTCC 1020  
 50 ATAACGATGA GTATCTGCTT CCGGAACCTT TTGGACACCT ATAACGAGT GCCCTGTTTC 1080  
 TTCATAAACG TCAATCAACT GTTTCACCTG TGGCACTTCA GATTCAACAA TATCGTCACC 1140

55

	TAAACCTTTT TGTTCTTTCT GCCTTACATA AAAAATATTC GCAAGTCCG TTGAATACTG	1260
	AACTTTCTCT AGTAATTCAG ATTTACCTTT TTCTTTTAAC ACCATTCTA ATTCTTTTGG	1320
5	ACTATCAAAA TGATCTTCAA TCGCGCGTTT GTGGCGACCT GTCACTATAA TAATATCTTC	1380
	AATTCCAGCT CTTGCAGCTT CTTCAACGAT ATATTGTATT GTGGGTTTAT CTAAGATAGG	1440
	AAGCATTTC TTTGGCATCG CTTTAGTTGC TGGTAAAAAT CTAGTCCCTA AACCAGCAGC	1500
10	GGGAATGATT GCCTTTTTTA TTTTTCCTAA AGTTAATGTG CTCCTTTTCC TAAGTATTAA	1560
	ATCTATGTAT CAACGTCATT TTAACACTAA TTAGAACGCC TTCATAGTGT CATTGAGTAT	1620
15	GTAATTATTT CTTGGGAAAT TTGTTTTAAT TTAAAAAAC AGGCTTACTT CATATAATTT	1680
	ATGAAATAAA CCTGTCAATT TTGGATTGAT TATGCTTTGT GATTCTTTT ATTCTGCGT	1740
	AATAACGCTA AACCTAAAAT GCTAAATAAT CCGCCGAACA ACATGCCGTT GTTTGTTGAT	1800
20	TCTTCTCCAC CTGTTTCAGG TAGTTCAGAT TTCTTAGATT GTGCTTTTTT AGTTGGTACC	1860
	ACTGCTTTAA CCTTTTCATT GATTTCAATA ACAGGTGTTA CTACTTTACC TTGTTCCACT	1920
	GGTTTAGAAG GTTTTTTAGG TTCTTCTTTA GCAGGTGGTA TTGGTTTACC AGGTTCAATT	1980
25	GGTACCTCTG GCGTTGGCGG TGTGGTGTT TCCGGCTCGC TTGGTACTTC TGGTGTCGGT	2040
	GGTGTGGTG TTTCCGGCTC GCTTGGTACT TCTGGTGTCG GTGGCGTTGG TGGCACGATT	2100
	GGAGGTGTTG TATCTTCTTC AATCGTTTGT TGACCTTCAT TATGACCACT TACTGTGGA	2160
30	AGTGTATCTT CTTCAAAGTC AACACTATTG TGTCCACCGA ATTGATAATT TGGTTTATCT	2220
	TTATTTGTAT CTTCTTCAAT AATTTCAAGT TGCTTATTGA ATCCGTGAAT ATGTGGCACA	2280
35	CTGTCGAAGT CGATATCAAT GATATTACCA CCTTGTTCAT ACTTAGGTTT GTCTTTCTCT	2340
	GTATCTTCTT CGAATGATTG GTTACCATTG TTTTGACCAT GAATTTGAGG TACACTATCG	2400
	AAATCGATAT CTACGATATT GCCACCTTGT TCATATTTTCG GTTTATCTTC TTCTGTGTCT	2460
40	TCCTCAAATG ACTGATTACC GCTATTTTGG CCACCTTCGT AACCTAATTC ACTCTTAATA	2520
	TCCACGTGGC TATTTTCTTC GATTTCTTCA ATCACGCCAT AATTACCGTG ACCATTTTCA	2580
	GTTCTTAAAC CAGAATGAGA AATATGATGA TTGTTTTTTCAG TAATTTCTTC GATTGGTCTT	2640
45	TGCGCTTGAC CATGTTCTTC AGGTAGTTCA TCTACTAGTT CAATCAGATT ACTTTCAGTC	2700
	GTATATTCTT TCGTATCTTC AATGTTGTA TGATCGCTAA CAGCACCAGT TACAATACCT	2760
	TTTGTAGAAT CTTTCGTCAAA TTCAACTAGG TTAGACTCAG TAGTAACCTG ACCACCACCT	2820
50	GGGTTTGTAT CTTCTTCATA TTCAACAACA TCAGCATGAT GTTTTGAATT TTCATGTGTC	2880
	GATTCTTCAA AGTCTACATG AATAGAATCT TCTTCAGTTT CAATGGTACC TTCTGCATGA	2940
55		

	TCTTCGATTG TACCAGTCAA TTCATGCTTC TCCACTGGCG GCTCTGATTT AAATTCAAGT	3060
	TCGATAGGAG TACTATGTTC TATAATAGGT TCCTTTAGTT TATCTTTGCC GTCGCCTTGA	3120
5	GCGTTATTAG AGTAAAATGC AACGCCATTT TTCCaAGTTA AATTACTTGT ATAATAATAG	3180
	TTATAATATC CAAAAAGGTG TGTTCGAAAT TCTAAGTTGC TAGCATTGA ATCATAATAC	3240
10	CCTTCATATT TTATTACATA ATTTTACTT TGGTCTAAAT TATTAAAGTT TAAAGAATAA	3300
	CCACCATTAG TATCAAAATC TAAACTCATA TTATCAGTCA CATCTTCAA TTTGCTGACA	3360
	TCATCAAGCT TTGCATAnTn AgctTTCAGC TAAATCGTCT GAACCAATGT GTTTATATAC	3420
15	CTTAACTGTT GGATTATTAA CCCCTGGTTT ATTTCTTTA GTTACTTGAC CAGTTACTGT	3480
	CACAGAGCTT AACGACTGGT TGTTAGGTTT CATGTACGCA AAATGACTAA ATTTCCCATC	3540
	TACTTTATTT AAAGTATCAA TTCGACCATT AGCTGTACT CCCCAATTAT CTCTAACTCC	3600
20	ACCTAAATAT TGAATATTAA ATATTTTGCT AACCGTAGTC TCACCCAATT TAACTTCAAC	3660
	ATTTTGGTTA CCTTTTTCG TCACTGTTGT AGGATCAATA AATAGATTTA AAGATAATTC	3720
	AGCAGTTAAA TCTTTCTTTT CTGTACATA TTCTTTAAAC GTATATCTAA CTTTCTTTT	3780
25	TCCAATTATT TCTCCTGTCG CCATAACTTG ACCATCTGTA CTTTTATCT CCGGAACTTT	3840
	ACGCAGTGTT GAGATACCAT GAGTTTCAAC ATTATCGCTT AATGTGAAAT CAAAATAATC	3900
	TCCCGCCTTA ATTCCTTCTC CAAATTTCCA TTTATATTTT AAGGTTACTC TTTCTGCGTT	3960
30	ATGAGGATTT ACAACATTCG TATCTTGTTT ATGTCCTACA ATTTCACTAC CTCTTCTAC	4020
	TTCCACTTTA TTGTGTACAT CTGTACCTGT CGCTTTAGTT TCTTCCACTA CTCTTTCTC	4080
35	TGCAACTGCT GTAACGTCAc TGatCTTTTC ATTCCTGGTT TAATTTCTGA GACGTTACTT	4140
	GGTTGAGCTA TGTCAACTTG AGTTCCTGTA GTTTCCTTAT CAGCAACTTT TTCCGATGGC	4200
	AAATCAACTC GCGAAgTTTC TACTTTTGGT GCTTGCAcAG TTTTCGGTGC TTCTTCTGTT	4260
40	GTTACTGTG TTGATTGTGA TGGTTGCTCA GTTGATGTCG CGCTGTATGA TTGTGTTTCA	4320
	TCTATTGTAT TAACGTTATT TGTAGTTGTT TGTGTTTCGC TTGCTTTACT TTCAGTAGCT	4380
	GAACCTCCAC TTTCTCTAC TGTAGTATTG TTTTGTCCG ATGCTGCAGC TTCTTTTCT	4440
45	TGTCCCATTC CAACAACGAT CATTGTTTCT AAGAATACTG AGGCCGCTCC CAATTGTGT	4500
	TTTCTTATGC CGTATCTAAG ATTGCTTTTC ACTATAATAT TCTCCCTTAA ATGCAAAATT	4560
	CATTTATTTT TAAACTCAA TAAATGCAAT TCTATATTGT TCGGTTTTTA AAAGCAATGA	4620
50	AAAAAAGCGA GTTAATAAAA AGTTAAGATT GTTGTTAACT TTATGTATAA TGAGTTTTTT	4680
	ATTATTTGAA ACTCACATAT ATATTGCATA CAAAGCTCTT GAACACCTTG ATATAACAGG	4740
55		

EP 0 786 519 A2

	TACTAAACCA TACATAATAA TCGCCTGTAC AATGCATCAT TAACAAGTCA CTGAAACGCC	4860
	TTTCATTGTA TTAATAACGT CACTATAATT TTTATATCGT TCGGTTTTTG TTTGATTTTA	4920
5	ATGATTATTT ATACAAAAC AGCCGTATTT CAAGCCGACA TTTTAAATTT AACTAAATTT	4980
	GCATCTAGTT AATAATTGCA TTTATCAAAT TTGTCTTATT GATCCAATCT AATTTGTACT	5040
10	CACAACTAG TTTAAAATTC TAACTTTATC TCTCAGTTTC TTATCAATCA TCAGACATAA	5100
	ACCAATGAAG CAATCAGAAA AACTCTAAT TTTCTATTAG AAATTTGATT TAATATAAAA	5160
	AAACAGGCTT ACTTCATATA ATTTATGAAA TAAACCCGTC AATTTTTGTT TAATTATGCT	5220
15	TTGTGATTCT TTTTATTTCT GCGTAATAAT GCTAAACCTA GAATGCTGAA TAATCCGCCG	5280
	AACAACATAC CTTTGTTTGT TGATTCTTCT CCACCTGTTT CAGGTAGTTC AGATTTCTTA	5340
	GATTGTGGTT TTTTAGTTGG TGCCACTGCT TTAACCTTTT CATTGATTTC AATAACAGGT	5400
20	GTTACTACTT TACCTTGTTT CACTGGTTTA GAAGGCTTTT TAGGTTCTTC TTTGGCAGGT	5460
	GGTACTGGTT TACCAGGTTT AGCTGGTACC TCTGGTGTG GCGGTGTTGG AGTTTCTGGC	5520
	TCACTCGGCA CTTCTGGTGT CGGTGGTGTG GGTGTTTCCG GCTCACTGG TACTTCTGGT	5580
25	GTTGGTGGCG TTGGTGTTC CGGCTCACTT GGTACTTCTG GTGTCGGTGG CGTTGGTGGC	5640
	ACGATTGGAG GTGTTGTATC TTCTTCAATC GTTTGTTGAC CTTCATTTTG GCCGCTTACT	5700
	TTTGAAGTG TATCTTCTTC AAAGTCAACA CTATTGTGTC CACCGAATTG ATAACCTGGT	5760
30	TTATCTTTAT TTGTATCTTC TTCAATAATT TCAGTGTGCT TATTGAATCC GTGAATATGT	5820
	GGCACACTGT CGAAGTCGAT ATCAATGATG TTACCGCCAT GTTCATACTT AGGTTTGTCT	5880
35	TTTTCTGTAT CTTCTCGAA TGA CTGATTA CTTTATTTT GACCATGAAT TTGAGGTACA	5940
	CTATCAAAAT CGATATCTAC GATATTGCCA CCTGTTCAT ATTTAGGTTT GTCTTCTTCT	6000
	GTGTCTTCCT CGAATGACTG GTTACCGCTA TTTTGGCCAC CTTCATAACC TAATTCATC	6060
40	TTAATATCAA CGTGGCTATT TTCTTCGATT TCTTCAATCA CGTCATAATT CCCGTGACCA	6120
	TTTTAGTTT CTAAACCAGA ATGAGAAATA TGATGATTGT TTTTAGTAAT TTCCTCGACT	6180
	GGTCCTTGTG CTTGACCATG CTCTTCAGGT AATTCATCCA CTAATTCAAT CAGATTACTT	6240
45	TCAGTTGTAT ATTCTTTTCG ATCTTCAACT GTTGTATGAT CGCTCACTGC GCCAGTTACA	6300
	ATACCTTTTG TAGACTCTTC GTCAAATTCA ACTAAGTTAG ACTCAGTAGT AACCTGACCA	6360
	CCACCTGGGT TTGTATCTTC TTCATATTCA ACAACATCAG CGTGATGTTT TGAATTTTCA	6420
50	TGTGTAGATT CTTCAAAGTC AATTGGATTT GATTCCTCAG AGGACTCAGT GTATCCTCCA	6480
	ACGTGACCTG CTTCGCTATC CACAGCAGTA TGGTAATCGA TATCAATAGC TGATGAATCC	6540
55		

	TGGTAATCAA TGTCAAGAGT TGATGAATCA TATTCCTCTT CAACAGTAGT TACTAAATTC	6660
	TTATCATATT GACCTGTAAG AGTTTCTTTA ATTGTATCTT CTTTATATTC AAATTTATTA	6720
5	TTTTGAATAA TCGGACCATT TTTCTCATTT CCGTTCGCTT TATTACTGTA TAAAACTAAA	6780
	CCATTATCCC AAGTTAAGGT ATATCCTCTA TCATAATAAT ACTTATAAAG TTGCTCTGGA	6840
10	TGTCCTACCA TTTGTGTTCT AAAATCAACT TCATCAGTAC CATTTAATA CTCTCCATCA	6900
	TAGTGAACAA CATAAGTTT ATCTAGATTT TCTATATTCA ATGAATAGCT TCCATTATTT	6960
	TGTAAATTCA AATTCCCACT CATATTACTT GTGACTTCTT TAAATTTAGA AGTATCTGTC	7020
15	GTATTTGCAT ATACACTCTT CGCTATGTCT TCATTATTAC CCAAGTATTC AAATATCCTA	7080
	ACTTTTGGTT GATTTCCATT CTGATTACTA CCTTTCATTA AAGTTCCAGT AACAGTCACA	7140
	CTTGTCGTTT TACCATTATT AGGTTTAATA AATGCAACAT GCGAAAATCT ATTATTCGCT	7200
20	TTATTAAATG TCTCAATCGA TCCATTAAAA TTGGCATAAT AATTCCCAAT ACCATCTTTA	7260
	TATTTAACAT CTAATTCCTT TGAAGTTTGT TCTTCATTTA GTGTTGAAGT TATAGTTTGA	7320
	TTTCCATTAG TTTGTACAGT TTTAGGATCA ATAAATAAAT TAATTTCTAG TTCAGCCGTT	7380
25	ACATCAACCT TATCTTCAAT ATCATTTGTA AATGTATATC TAATCTTTCC ACCTTCTAAA	7440
	ACTTCACCTG TCGCCATTAC GACTGAACCA TTTTAAATTT CTGGTACTTT TCTAGCAGTT	7500
	GATACGCCAT GCGTATTTAC ATTATTTGAT AAAGTAAAGT CAAAGTAGTC ACCTTGATGT	7560
30	AAACCATTCT CAAATTTCAA CTTATATTTT AGTACCGCTC GTTGTCTGTC ATGAGGTTCT	7620
	ACTTTATTTG TATTGTTATG CCCCTCAATA GAACCAATTT CTACTGTAAC TTTACTTGTT	7680
35	ACATCTGTAC CCGTTTCCAC TTTCGCGTTA CTAGCTTCCT TAGCTTCCGC TACATCTGCT	7740
	GATCTTGTC ACGTGGCTT ACTTTCTGAT GCCGTCTTG GCTGTGCCAC TTCAACTTGT	7800
	GTTTCTGCGA CTTGATTTTG TGTAGCCTTT TTAGGTGTTA AATCTACTTG TCTTTGATCT	7860
40	CCGCTATTGT CTTGAGATTG TGTGTTTCC TTAAGTTGAG GTTTCGCTTC TTCCTTAACT	7920
	ACCTCTTCTT TAACTGTTTC TATATTTGCT GGTGTGCAG TTTGTGGTGC TTGTACTGCT	7980
	TTTGGTGCTT CTTCAAGTTGT TACTTGTGTT GCGTTTGACG GTTGTCTGT TACTGTTGCG	8040
45	TTATATGATT GAGTTTCTTC TATATGATTA ACGTTAGTTG CAGTTGTTTG TGTTTCACTT	8100
	GTTTATTAT CAGTAGCTGA ATTCCCATTT TCTTCTACTG TAGTTGTCTT TTGTTCTGAT	8160
	GCTGCAGCTT CTTTGTCTTG TCCATCCCA ACAACGATCA TTGTTCTTAA GAATACTGAT	8220
50	GCTGCTCCCA ATTTATGTTT TCTAATGCCG TACCTAAGAT TGTTTTTCAC TATAATATCT	8280
	CCCTTTAAAT GCAAAATTCA TTAATTTTIT AACTTAATA AATGCAAGTC TATATTGTTT	8340
55		



EP 0 786 519 A2

	ATGTTAATTG ATAATTTTAT TATTTGAAAT ATACCTATAA ATTGTATTCA AGTCATCAGA	8460
	AACCCCTTGTC ACACAAGGCT TGTATTTTTT ATACTTATTT TTTAAATTAA ATTTCATCATT	8520
5	ATCTAATTTA AAACAATATA CTAAACGTTT CATAATTATC GCCTGTACAA TACGCACAAA	8580
	AACATGTCTT GAAACGCCTT TCATTACTCT AAAATACCCA ATATACTTTT TATATCGTTC	8640
	GGATTCTGAG TATTTTCAGAC GATTTTCTGC ATAAAAATAA ACGTGTTCCTA AGGCAATATA	8700
10	TTGCAATTAC CTAAAAACAC GTTTACTTAA TATTTAGTTA AACAAATAAG CTAATGAATA	8760
	AAATGAAGAT GATACCTGAA ACGGAAATAA TCGTTTCTAA TAATGACCAT GTTAAGAATG	8820
15	TTTCTTTTAC AGTTAAACCA AAATATTCTT TAAACATCCA AAATCCTGCG TCATTTCAT	8880
	GAGACAAAAT CACACTACCT GCACCTATCG CAAGTACAAC TAATGCAACA TTTACATCTG	8940
	ATGATTGTAA TAATGGTAAG ACAATACCTG TAGTTGAAAT CGCAGCTACT GTAGCCGAAC	9000
20	CTAATGCGAT ACGTAGCACA GCTGCAACAA TCCATGCTAG TAAAATCGGA GACATCTCTG	9060
	TACCTTCAAA CATTTTAGCA ATTGTATTTT CGACACCGCC GTCAATTAAT ACTTGTTTAA	9120
	ATGTACCGCC ACCGCCAATA ATCAATAACA TCATTCCGAT TGGATAAATC GCATTCTGCA	9180
25	CTGATTCCAT AATATGATTC ATCTTACGCT TTCTCATTAA TCCCATCGTA ACGATTGCAA	9240
	ATAATACTGC TATTAGCATG GCTGTCCCTG CTGTTCTTAT CATATAAATG ATAGATTCAA	9300
	ATAGATTGTG AGGTTTGTCA TGCCCAGTTA CAAGTTGCGT TATCGTAGAC ACTAACATTA	9360
30	ATATGACTGG TAATGTTGCT GTTAATAAAC TCATACCAA TCCTGGCATC TCTTGATCCG	9420
	TAAATTCTTT TTGTGCACCT AACGCTGAAA TATCGCCTTC TCGTGATAC GCAGACGGAA	9480
35	TCATTTTTTG TGCACTTTGT TAAATATAGG CCCTGCAATG AGTGTAATG GAAATGGCAAT	9540
	AATCATACCA TACAGTAATA CATCTCCAAC ATTTGCCTTT AATTCTTTTG CGATGACTAC	9600
	CGGTCTCGGA TGTGGTGGTA AAAAGCCATG TGTCACTGAT AAAGCTGTTA CCATAGGTAG	9660
40	TCCTAGTTTT AACACTGAAA CATTTGCGCG TTTTGCTACT GTAAATACTA ATGGAATCAG	9720
	TAAGACTAAA CCTACTTCAA AGAACAATGC AATACCGACG ATAAATGCTG CAACAAGCAT	9780
	TGCCCATTGT ACATGTTTTT GACCAAATTT TTGAATCAAC GTGTCTGCGA TTCGAGTTGC	9840
45	ACCACCACCA TCAGCAAGCA ATTTCCCAAG TATGGCACCT AAACCGAATA TCAGTGCAAT	9900
	GTGGCCGAGC GTACTGCCCA TTCCTTTCTC AATCGTCTCC ATAATTTTAG TCAATGGTAT	9960
	ACCTAGCATT AACGCTGTAA TCATCGATGT GATAATTAAT GAAATAAATG TATTTAATTT	10020
50	AAACCCAATA ATTAATACTA ATAAAATAAC GATACCTAAA ACAACACTGA TTAACGGCCA	10080
	TATTTCTGTT AACATGACAT TCCCCTCTTT CTCTTTTCAA TAGAATGTAA CACCGTCGTC	10140
55		

	GAGTGACGTA	TTTATTGTGT	TTTATTTTCA	GCGATATGTT	GGCGTTGAAA	ATCTGCAATT	10260
	TGTTTCATAAT	TCTCTGTAA	AGAACGACTT	AAATTGATAA	AAATGGATAC	GATCTCTTGG	10320
5	TAAACAGTGA	CATTTTCTTC	AATCGGCGTA	TGATTGTTTG	TGGCACCGAC	CATCGATGAA	10380
	ACGATTGAAA	AATCTTCAAT	GTCACCTACA	GCTTTAAGTC	CGAGCACGCA	GGCACCTAAG	10440
10	CATGAACCTT	CATAACTTTC	AGGAACCACT	AACTCTGTGT	CAAATATATC	TGACATCATT	10500
	TGACGCCATA	CTTCACTTTT	CGCAAAACCA	CCTGTTGCTT	TTATCATCTT	AGGTGTTTCA	10560
	TTCATTACTT	CAATAAGCGC	AAGATAGACG	GTATACAAAT	TGTAAAGAAC	ACCTTCTAAT	10620
15	GCAGCGCGAA	TCATATGTTT	TTTTTTATGA	GATAAAGTTA	AACCGAAGAA	TGAACCTCTT	10680
	GCATTTGCGT	TCCAAAGCGG	CGCACGTTCT	CCTGCTAAAT	AGGGATGGAA	TATTAAACCA	10740
	TCTGCACCTG	GTTTAAACAG	CTTTGCAATT	TGAGTTAAGA	CATCATAAGG	ATCAACACCG	10800
20	AGACGTTTCG	CAGTTTCGAC	TTCACTCGCT	AGCAACTCGT	CGCGCAACCA	TCTCAATACG	10860
	ACACCACCAT	TATTTACAGG	ACCTCCGATG	ACGTAGTGGT	CCTCTGTAA	GACATAACAA	10920
	AATATTCTAC	CTTTGTAATC	AGTACGCGGT	TTATCTATCA	CAGTACGAAT	CGCCCCAGAT	10980
25	GTACCGATTG	TGACAGCAAC	TTCTCCTTTA	CCAACACTAT	TGACACCTAA	ATTAGAAAGG	11040
	ACCCCATCAC	TCGCACCAAT	AACAAACGGT	GTATCTTTAT	TAAGCCCCAT	TAATGTTGCA	11100
	TAACGTTCTT	TCATACCTTT	CATCACATAC	GTTGTTGGAA	CTAATTCCGG	CAACATTTCC	11160
30	TTGGAAATAC	CCAGCAGTTC	TAATGCCTCA	ACATCCCAAT	CTAATGTTTC	TAAATTAAAC	11220
	ATCCCTGTTG	CGGAAGCCAT	TGAATAATCA	ATGATATATG	TATCAAATAA	ATGATAGAAA	11280
35	ATGTATGTTT	TAATATCTGC	AACTTAGCA	GTACGTTGAA	ATACATCTTG	CCATTCATGT	11340
	TTCATCCAAA	AAATCTTCGC	TAATGGCGAC	ATAGGATGAA	TCGGTGTGCC	TGTTGCTGCTG	11400
	TAAATCGCAT	TGCCATCATG	CACTTCATTT	ATTACTGTTG	CATATTTTGC	AGCGCGGTTA	11460
40	TCTGCCCCAAG	TAATATTATT	TGTTAATCTT	TGATGTTGCT	GATCCATCGC	AATCAAGCTA	11520
	TGCATTTGCG	CACTAAATGA	CACAACTTA	ATGTCGTCTT	TATTAACCTT	GGATTCTCTC	11580
	ATAACATATT	TAATAGTCAT	TAGTACTGCA	TCAAATAATT	CATCTGGGTT	TTCTTCTGAG	11640
45	ACATCAACGT	TTGGTGTGTG	TAAATCATAG	CCTATTTGAT	GTTTCATGAT	AAAAGTTCCA	11700
	TTTTCATCAT	ATAAGACTGA	CTTGGTACTC	GTCGTTCCAA	TGTCGACACC	AATCATATAT	11760
	TTTCATGATAA	ATCCTTCTTT	CTTTCATTTT	AATTCAACCA	AAATCCTTCA	ATATCTTTAC	11820
50	CAACATCGTC	GAAATTTAAA	TGAAACGCTT	CTTTCAAAAT	TTGACTGTCG	TATTGTTCCA	11880
	CTGCATCAAT	AAACACTTGA	TGATTATGAT	GTATGCGTTC	AAAATCTTGC	GGGTTCTGTT	11940
55							

	AAAATGAGTT TAAATATTGA TGATTAGATG CTTTGATTAA TGTTTCATGA AATTCAAAGT	12060
	CATGCTTCGT AAATGATTCT GCATCCTCAA ATTTTACTGC CACTTTCATC ATTTCAAGTT	12120
5	GTTTCTTCAT TTCTTTTACG ATAGGTAGTC GCTCTTGATT TTAACTCTT GAAAATGCAA	12180
	ATGACTCTAA CATCAGTCGC AAATCATACA TTTCTTTCTT TTCTTGTTC CCAAACGGCA	12240
10	ACACATGTGC ACCCATTTCTT TCTAATTGGA TGAGTTGATT TTGTTGCAAT AATTTAAATG	12300
	CATCTCGAAT TGGCGAACGA CTCACATTAA ATTGCTTTGC CATTTGATTT TCAGTGAGTA	12360
	ACGTACCTTC AGCTATGTGA CCATTCACAA TGCCTAAGCG TAATTCTGCC GCGATACCTT	12420
15	CTCCAGTTGT CATACCTTCC AACCATTTCT CTGGATATCC ATACATCATC AAAGTCACTC	12480
	CTTCATTACA CGACATACTT GTATACAAGT ATGTTAATAT AGTTATTATG AGTTTGCAAG	12540
	CGCTTTCTTT ACGAGCACTA AAATAGTGAC CACCCCTTTT CGATTTAAAT TTAAAGGAAA	12600
20	TGGTCACTAT CACACGAATG ATTTAATTGT TATGTTGTAT GTGGGATATT TCTAATTGTT	12660
	CTGTACTCAT ATGCGCTTTA GGTACTTCAA TGCAATAATG CGTTTCATGA CAGTTTGGAC	12720
	ATTCGAATCG ACGTGTGTGC GCTGTATGTT TCGCTTTGAT AACTGCCCAC AAAGATGGTG	12780
25	AGAATATATG CTGGCAGTTA GGACATAAAT AGGCAACCTT TTGTTGGTAA TAAAAAGTAA	12840
	CACCAATGCC ATAACCAATC ATAAATGGTA AAGCAATTAA AAACGGCCAT TTATTTTTC	12900
	TCAAAATTGC ACTTATAATG CTAGAATATT GAATTATTCC TATAATACCA GCACTAATCC	12960
30	AAATGTTACG ACGAATACTT TTCATTTTCA CTGATTTACT CATGACATGC TCTATGTCTT	13020
	TTAAGTGTGT GATTGGAGAC GTCGACGCTT CATTTACGTA ATATTGAACA TTTTAAATTT	13080
35	TGTTTAATAC CGCTTGTGTC TGTTTAACTT GTTGGTTAAT TTCTTGTGTG TTCATAGTTA	13140
	GTAAAGTATT GAGCGTCTTC AAAGTACCTT CACCTTTTAG CAACATATCT ATATCGCTTA	13200
	ACGCAACAACC TAAATCTTTA AGCAATAAGA TTAACCTCTAA TGTTGTGCGC TGTTGTTCTG	13260
40	TATACACACG ACGCTTTCCT TCTGTAAATC CTTGTGGTTT CAAAATACCT TTGCGATCAT	13320
	AATATTGAAT CGTTCGTGTT GTCACATTGC ATAATTTTGC GAGTCTCTCA GTCGAATAGT	13380
	TAGACATAGA TTCCACCTCC TATAATTACC ATAGTTGATG ACCCGACGTC ACGAGCAAGT	13440
45	ACAATTTCCA CATTTTAAAG AAATTTATTA TACTAGGCGT CTTATTTTAA TGATTTTCGTA	13500
	CCATGTTGAT TTACAAACTC ACTCAAATA AGTAACACAC CTACTAAACA TCTACTCTGT	13560
	TATTTAGAA TGAATTTGTT GTAATTTATC TTCAACTTCA GTAATCTCTG TCGCACATTC	13620
50	TTTCAGTAAA TCTCGATACT TTTCCGTCTC TGCAATTGTT TTATAACGTA TTTTATGTTT	13680
	TAAACTTGcC CACATATCCA TACCTATCGT TCTAATTTGA ATTTCAACAG GCAATACCTC	13740
55		

## (2) INFORMATION FOR SEQ ID NO: 55:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1059 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

```

GGATAAGTTC AGGTAAATTC ATTTCTTTTT CAATTTTGAT TTTCATTGTT TCCGCCCTTT      60
TAAAATAAAG TTAGTTGCTT CTGTTCTCTCA TATTCCAAAT CACTTTGCTT TATATATGTT      120
TCAAGCTCTT CCGCTGTATC AAATGTCCTT TTCACACCTT GCCAACCTGG CACGATATGA      180
CCGTGAAAGT AATAAGTGCC ATTTACTACA TGGATATGTG CCACTCGTTC GTTATCCTGA      240
TACAGATATC TCTTAGATCC AAAGAATTGA TTTAGGTATT CTTTACGCGC GCTATCTGTC      300
ATGGTCATCA CTCCTTTTAA CAATTAGGCA GACCAAACGA CATGCATTCTG TCGTATAGCT      360
CTTCATTACT TATGCTTGCC TTATAGTTTT CAATCACATT GCTAACTTCT TTATGACTCA      420
TTGCTTTAAC TTGTTCTGCT GTATATTTTT CGCAGTCTTC TAATTCCAGT TGCTCCTGTA      480
ATGACATCAC ATATTCAACT TGTCTTTGGG TTGCCATCGT TAACCCTCCC ACAAGTCAAA      540
AGCTCTTTGG ACGTAAAACT TCGCCTTTGC TAAATCCTCA TGACCATTCT TTAACGGTGC      600
TCTAGACATG TATTTGATTG CATTACCTAT TGCGAATGCT AGTTGAGGTG GATACTGTGC      660
CGTAACCTGT TCGATAAAAT CTATAATTTC AATGTCGCCG TATGTGTAGT GCGCTGGTTG      720
CTTAACATTG TCTTGCGCTT CGTTCATATC TACTTTTCTG TTACTGATTA CGCTCATTAT      780
GCTTCACTCC ATTCTTGAA CATTTGGTTA TAAGTGACAT CGAACCAGTA CGGATCACGT      840
GAATGTTTTT GTGGCGTTCC ATCATAAAGC CATGGTCTTA ATCTTCTCTT TCTTTCCTGT      900
TCATATTCCG CTCTCACATT TCGTTGGTAT CGGTTCAAAA TCGCTTTTTT TCTGATTTTT      960
TCTCTCCCTT TTTCTTCATC TTTnATtTGA CTCTnCATAT ATTCAACTTC TTCTGTAGAT     1020
nTTGAGTCCT TTCTTCCACA CAATAATTCA nCGCCGCGC                               1059

```

## (2) INFORMATION FOR SEQ ID NO: 56:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30246 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## EP 0 786 519 A2

	GAAGTAAAAG AAGAATTAAA TTTAACATTA ACAATGGATG AAATTGAATA TGTCGGGACA	60
	ATTGTAGGTC CTGCATATCC ACAACAGGAT ATGTTAACTG AGTTAAATGG ATTTTCGCGCA	120
5	TTAACCAAAA TCGATTGGGA AAACGTAACT ATCAATAATG AAATTACGGA TATACGCTGG	180
	ATTGATAAAG ATAATGATGC GTTGATTGCG CCTGCTGTCA AAGTTTGGAT TGAAACTTAT	240
10	GGTGGTAAAC ATGACAAATA ATGACACCAT CATGTTACGA CATTATGTCC CACAAGATTA	300
	TTCGATGTTA GAAGCTTTTC AATTAAGTGA AAGTGATTTG AAGTTTGTTA AAACGCCAGA	360
	GGAAAATATT ACAGCTGCAA TGTCTGATAA TGAAAGGTAT CCCATCGTTG TAATGGATGG	420
15	CAGGCAATGT GTGGCCTTTT TTACATTACA TCGTGAAAAA GGGGTGCGAC CATTTAGCGA	480
	TAACCAAGAT GCAGTATTTT TCAGGTCATT TAGTGTGTAT CAACGTTATC GTAATAGAGG	540
	AATAGGTAAA GTGGTAATGG AAAAATTGGC GTCATTTATC ACTTCAACAT TTCAGGATAT	600
20	TAATGAGATT GTGTTAACGG TTAATACTGA CAATCCACAT GCCATGGCAC TTTATCGCCA	660
	ACAAGGATAT CAATATATGG GAGATAGTAT GTTCGTCGGA AGACCTGTTC ATATTATGGC	720
	GTAACTATA AAATAAATTA AATTTAAAAG CATCTTTACT CATCGTCGAC CACAACAATT	780
25	AATGATGAAT AAAGGTGCTT TTTGTTATAG ATCATCGGAC AATTTACTAT AGTAAAAAGC	840
	GACCTAGTGA ACAATTGACA TATATCCACA GGTGCTTAA CTTAAGTTAT ATTGCTAGTT	900
	GCGATTAATT GATAGACTCA TCATTTTTCG GCTGTCGAGA TGGTCTTTTT ATTAATAATG	960
30	CCGTAATCCA AGCCGTAATC GGAATACTGA TTGCAACGGC AATACCGCCT AAAATAATAG	1020
	AAATAAATTC TTGGGCAAAT ATTTTCGAGT TTATAATATG ACCAAATGAA TATTTAAGTT	1080
35	TGAAAAACCA AATAAATAAA GCAAGTTGGC CACCAAAAAA GGCAAGGTAA ATCGTGTTCG	1140
	CAGATGTCGC TAAAATTTCT CTACCAACAC GCATGCCAGA TTGGAATAAT TCGTATTGCG	1200
	TAACGTTgGA TTCACTTGAT GCAATTCATA AATGGGTGAA CTAATGGTAA TTGTTAAATC	1260
40	TATCAGAGCT GCAATAACAG CAAGAATAAT AGTGAACACC ATAAATTGAA CCATATCAAT	1320
	GCCAATATTC ATTGAATACA CATATGTTTC ATCTTGTGTG TCGGTTGaAA AGCCTTGTAG	1380
	ATGACCGAAG TAGACCGATA AATAAATGAG TGTAATCAAC AATATTGTTG TAACGATAgT	1440
45	GCTGgATAAA TGCaGCTTGT GTTTTAAACAT TGTAATATT GAGTACGAAT AAATTACAAG	1500
	CGCCAATAAT AATGCAGAAA AAGAATGTGA CGACATAAAT CGGTACGCCA AAAATAATCA	1560
	ATACAATACT AATAATTAAA ATAGCGAAAT TTAAAAATAG GGTAAATAA GAGATGAATC	1620
50	CCTTTTACC TCCGAAAATT ATCATCAGAA AGAGGAGCAA TAACGCCAAT ATAAATACAG	1680
	CATTCATTGT TTCGCCCTCC TTAATGTTTC AAATATTTCC ATAAACAATA TTGTGATAGG	1740
55		

EP 0 786 519 A2

	CATCGAAATA GTATAAGTCA CTGTATTGGC ATTTTTTTAA AAGATTAAAA ACATAGGTAG	1860
	TGCACCGGAT AAATATGAGA ATAATAAGAT GTTACTCATT GTTCCCATAA TATCTTGGCC	1920
5	GATGTTTCGC CCAGCAAGCG CCCATCTCCT CATTGAAATG TGTGGCGTAC GCTGTAAAAT	1980
	TTCATGCATA CCACTAGCAA TTGTAATTGC AACATCCATA ATAGCGCCAA GTGAACCTAT	2040
	TAACACTGAG GCTAGGAAGA TATCTTTCGG TGGTAATGAT AAAAAGTTCA TCGTTTCATA	2100
10	TTTAATGCCT TTACCATCTG TCATATATAT GATTAATTCT GTTAAACCTA TACTCAAAAA	2160
	AGTTCCGATA ATTGTACTGG CTATGGTAAT GAGTGACGC ATATGCCAGC CTGTAACGAG	2220
15	CAATAAAGTG AGTATTGTTG AACAGATCAT GGCAATGGTC ATGAGTAAGA ATAAATTAAT	2280
	ATTGCTATGT TGAATATGAA TGTAATTGC GATTAATATG GCAATAGAAT TCAAGATTAA	2340
	CGATAAAATC GATTGCAGTC CGACTTTGCG ACCAACCAAT AATACAGTTA ATAAGAACAA	2400
20	ACCAGTGATG ATAACCGTTA AGGTATCACG CTTCTTTTCT ATAATATAAG CATCACTCGG	2460
	CTTGTTAGAA ATATGTAATA ATACTTTTTC GTGTGTGCGA AATGCCTCAG AATCTGCTTG	2520
	CGATTTGACG TACTGATGAT TAATCGTCGT CGTTTCTCCA GCAAATTGAC CATTTAATAT	2580
25	TTTGACTTTT AATTGATTTT TATATTTAAT ATCACGATTA TTTTGTGCAT CTTTGTAGG	2640
	TGTCGAAGAA ACATGTTTGA CATCTATAAT TTGACCAATT GGTTTGTTGT AAAAGTTCTC	2700
	ATTATTGAAT GTAAATAAAA TAGCACCAAT GAATGCGATG CAGAACAAAC CTAAAATTAT	2760
30	ATTAAATGGC TTTGTAAATA AATTTCTATA TTTCAAAAAC AAAACCCCAA TTCTATGAAT	2820
	GAATTAATAT GGTGATTATA CGCCCTTAAT TTTTATTTT CAAAGATATT ACTGCTAAGT	2880
	GTAAAACGAA AATCATCATT GATAGCATCG AATTACTTAA TGGAATGTAG ACGTTTTAGT	2940
35	CATTAATTGC TGAATAAGTG TTAATAATAT GCCAATATCA CTCTTTGTAT AAGGCTCCTT	3000
	TGTAATAGCA CATATCGTTC TTTTAAATTC AGTATGATCT AATTTTATAT CTATCCATGA	3060
40	TTTÀGATTCT GGTAAATGTA TATTTTGTGA TGAAATGATG TAACCTTCTT TTTGACGAAG	3120
	GAGATAcTGC GCAAGTGGTT GGCTACTGAT TGTGTATACA TCTGATTAG TAATCTTGCG	3180
	CAATTGTTTT TTTACAGTTT CGGCAAATGG TGCCAAGCAA TAAATATGAC TATGCTCAAA	3240
45	CTGAATTAAT GGTGGGTGTG TCGCCATCGT AATTGGATCG TCTGAAGGCG CATATAAATG	3300
	ATAGTGCTCT TCGAATAAAG GTAGCATATG TAATTGTTTG TGTTTACGTA TTTCTGGTGT	3360
	AAGTTCCGTG AAACCAATGT CTATATTCCC ATTTAATACG CTATTTATAA TTGTGTCATG	3420
50	TTCTAATAAG CTCGGTATGA CATGTGTATC ATTTTGTAAG TGAAACGTTT GGATAAGTGG	3480
	TAGTAACATG TGGGATACGT CACTCTCATC ATAGCCAATG TAGATACTTT TATTTTTAGT	3540

55

	TTCATTAAAT AATAATTTCC CTTCAGATGT GAGCGTAATA TTGCGTCCTT GCTTTTAA	3660
	TAAAGACACA TTAAGTTCTT GTTCTAATAA TGTAATTTGA CGGCTTATCG CTGATTGAGC	3720
5	AATGTTTAGT TCAAGTGCTG TTTCCGAGAT ATGTTCTCTT TTAGCGACCT CGATAAAATA	3780
	TCTTAATTGT TTAATTTCCA TAGCGATATA GGCACCTCCA AAAATGAGTG TTTTGTA	3840
	ATTATAGCAA TATTATTGAT AAATGTTCTA TTTTITAGAT GAATATCTTC TATTTTATAT	3900
10	ATTGAACAGA TAAATTTTTT AGATTATAGT AATTATCATT AATAACTAAT ATCAGAATAT	3960
	TCTAAAAAAG GGGTGTGCAT CATGCACAAT GAGAAATTAA TTAAAGGCTT ATATGACTAT	4020
15	CGTGAGGAAC ATGATGCGTG TGGTATTGGT TTTTATGCGA ATATGGATAA TAAAGGTCT	4080
	CACGACATCA TTGATAAATC GCTTGAAATG TTGCGACGCT TAGATCACAG GGGCGGGGTC	4140
	GGCGCAGATG GCATCACTGG TGATGGCGCA GGTATTATGA CTGAAATACC TTTTGCA	4200
20	TTCAAACAAC ATGTAACGGA CTTTGATATC CCAGGTGAAG GTGAATATGC CGTGGGGTTA	4260
	TTTTTTTCCA AAGAACGCAT TTTAGGTTCT GAACATGAAG TAGTTTTTAA AAAATATTTT	4320
	GAAGGCGAAG GGTATCAAT TCTTGGTTAT CGTAATGTAC CAGTTAATAA AGATGCCATT	4380
25	GCTAAACATG TAGCAGATAC GATGCCAGTC ATTCAACAAG TGTTTATTGA TATTAGGAC	4440
	ATTGAAGATG TTGAAAAGCG TTTGTTTTTA GCGAGAAAAC AATTAGAGTT CTATTCGACT	4500
	CAGTGCGATT TAGAATTGTA TTTTACGAGC TTATCACGCA AAACAATTGT ATATAAAGGT	4560
30	TGGTTACGAT CAGACCAAAT TAAAAACTA TATACAGATT TATCGGATGA TTTATATCAA	4620
	TCAAAGCTAG GGTTAGTGCA TTCGAGATTT AGTACGAATA CATTCCTGAG TTGAAAAGG	4680
	GCACATCCTA ACCGTATGTT AATGCATAAT GGTGAGATTA ACACGATTAA AGGTAATGTA	4740
35	AACTGGATGC GAGCACGCCA ACATAAATTA ATCGAAACAT TATTTGGCGA GGATCAACAT	4800
	AAAGTGTTTC AAATTGTCGA TGAGGATGGT AGTGACTCTG CCATTGTAGA TAATGCGCTA	4860
40	GAGTTCTTAT CGTTAGCCAT GGAGCCAGAA AAGGCAGCGA TGTTACTCAT ACCTGAACCT	4920
	TGGTTATATA ATGAAGCGAA TGATGCAAAT GTACGTGCGT TTTATGAATT TTATAGTTAT	4980
	TTAATGGAAC CGTGGGATGG TCCTACAATG ATTTCTGTTCT GTAACGGTGA CAAACTTGGC	5040
45	GCGCTTACAG ATAGAAATGG ATTACGTCCA GGTGTTTATA CGATTACTAA AGATACTTT	5100
	ATTGTCTTTT CATCTGAAGT GGGTGTGTG GACGTACCTG AAAGTAATGT TGCTTTTAAA	5160
	GGTCAATTGA ATCCTGGAAT GTTATTGCTT GTTGATTTTA AACAGAATAA AGTCATTGAA	5220
50	AATAATGATT TAAAGGTGC GATTGCTGGA GAATTACCAT ATAAAGCGTG GATTGATAAC	5280
	CATAAAGTTG ACTTTGATTT TGAAAATATA CAATATCAAG ATTCGCAATG GAAAGATGAG	5340

55

	CAGGAACTTG TAGAAGGTAA GAAGGATCCT ATCGGTGCAA TGGGATATGA TCGGCCAATT	5460
5	GCACTGTTGA ACGAGCGACC AGAATCACTA TTTAATTACT TTAAACAGCT GTTTGCACAA	5520
	GTTACGAATC CACCAATTGA TCGGTATCGT GAAAAAATCG TAACGAGTGA ACTTTCTTAT	5580
	TTAGGTGGCG AAGGTAACCT ACTAGCACCT GACGAAACGG TTTTAGATCG TATTCAATTG	5640
10	AAAAGGCCCG TATTGAATGA ATCACACTTA GCAGCGATTG ATCAGGAACA TTTTAAATTA	5700
	ACTTATTTAT CAACGGTATA TGAAGGGGAT TTGGAAGATG CGTTAGAAGC ATTAGGCCGA	5760
	GAAGCAGTGA ATGCTGTAAA GCAAGGCGCT CAAATTCTAG TGTTAGATGA TAGTGGATTA	5820
15	GTTGATAGCA ATGGCTTTGC AATGCCGATG TTAAGTCATGT GCATCAATTA	5880
	CTTATTAAAG CAGATTTACG TATGTCTACA AGTTTAGTCG CTAAATCTGG TGAGACACGA	5940
	GAAGTGCATC ATGTTGCTTG TTTACTCGCA TATGGCGCGA ATGCAATTGT GCCATACCTA	6000
20	GCGCAACGTA CAGTTGAACA ACTGACATTG ACAGAAGGGT TACAAGGCAC CGTTGTCGAT	6060
	AATGTTAAGA CATATACGGA TGTATTGTCA GAAGGTGTCA TTAAAGTAAT GGCTAAGATG	6120
	GGAATTTCTGA CAGTGCAAAG TTATCAAGGG GCACAAATAT TTGAAGCGAT TGGCTTGTCT	6180
25	CATGATGTGA TTGATCGTTA TTTTACTGGG ACACAGTCTA AGTTATCTGG TATTTCTGATT	6240
	GATCAAATTG ATGCTGAAAA TAAAGCACGT CAACAAAGTG ATGATAATTA TCTTGCATCA	6300
	GGTAGTACAT TCCAATGGAG ACAACAAGGT CAACATCATG CTTTTAATCC GGAATCTATT	6360
30	TTCTTATTGC AGCACGCATG TAAAGAAAT GACTATGCGC AATTTAAAGC ATACTCTGAA	6420
	GCGGTGAACA AAAATAGAAC AGATCACATT AGACATTTAC TTGAATTTAA AGCATGTACA	6480
	CCGATTGACA TCGACCAAGT TGAACCGGTA AGTGACATTG TCAAACGCTT TAATACAGGG	6540
35	GCGATGAGTT ATGGATCGAT TTCAGCGGAA GCACATGAAA CGTTAGCACA AGCCATGAAC	6600
	CAATTAGGTG GAAAGAGTAA TAGTGGTGAA GGTGGCGAAG ATGCAAAACG TTATGAAGTA	6660
40	CAAGTTGATG GAAGCAACAA AGTAAGTGGC ATTAAACAAG TTGCTTCTGG GCGTTTTGGT	6720
	GTAAGTAGTG ATTTATTACA ACATGCCAAA GAAATTCAAA TTAAAGTTGC GCAAGGTGCA	6780
	AAGCCTGGTG AAGGTGGTCA ATTACCTGGT ACTAAGGTAT ATCCGTGGAT TGCGAAGACA	6840
45	AGAGGGTCAA CGCCAGGTAT CGGTCTGATT TCACCACCGC CACATCATGA TATTTATTCA	6900
	ATAGAAGATT TAGCGCAACT GATACATGAT TTGAAAAATG CGAATAAAGA TGCAGATATC	6960
	GCGGTAAAAT TAGTTTCGAA AACAGGTGTT GGTACCATG CATCTGGGGT GGCAAAAGCA	7020
50	TTTGCAGATA AAATTGTCAT CAGTGGTTAC GATGGTGGTA CAGGGGCTTC ACCCAAAACG	7080
	AGTATTCAGC ATGCCGGTGT TCCTTGGGAG ATTGGTTTAG CAGAAACACA TCAAACATTA	7140

55



	AAAGATGTAG	CGTACGCATG	TGCGCTTGGA	GCGGAAGAAT	TTGGATTTGC	AACTGCACCA	7260
	TTAGTGGTGT	TGGGCTGTAT	TATGATGCGT	GTATGCCATA	AAGATACATG	TCCAGTAGGA	7320
5	GTTGCAACTC	AAAACAAAGA	TTTACGTGCT	TTATATAGAG	GTAAAGCACA	TCATGTTGTT	7380
	AATTTTATGC	ATTTTATTGC	ACAAGAATTA	AGAGAAATTT	TAGCATCTTT	AGGTTTGAAA	7440
	CGTGTAGAAG	ACTTAGTTGG	AAGAACTGAT	TTATTACAAC	GATCATCAAC	ATTAAAAGCG	7500
10	AATAGCAAAG	CGGCTAGTAT	TGATGTTGAA	AAACTGTTAT	GTCCTTTCGA	TGGGCCAAAC	7560
	ACAAAAGAAA	TTCAACAAAA	TCATAATCTT	GAGCATGGAT	TTGATTTAAC	AAATTTATAT	7620
	GAAGTAACGA	AGCCATATAT	TGCTGAAGGG	CGTCGCTATA	CAGGTAGCTT	TACAGTAAAT	7680
15	AATGAACAAC	GTGATGTAGG	GGTTATTACA	GGTAGTGAGA	TTTCGAAACA	ATATGGAGAA	7740
	GCAGGACTTC	CTGAAAATAC	AATTAATGTT	TATACGAATG	GTCATGCTGG	TCAAAGTCTT	7800
20	GCAGCATATG	CACCGAAAGG	CTTAATGATT	CATCATACTG	GAGATGCGAA	TGACTATGTT	7860
	GGTAAAGGAT	TATCTGGTGG	TACGGTCATT	GTCAAAGCAC	CTTTTGAAGA	ACGACAAAAT	7920
	GAAATTATTG	CTGGTAACGT	CTCATTCTAT	GGTGCACAA	GTGGTAAGGC	ATTTATTAAAC	7980
25	GGTAGTGCAG	GAGAAAGATT	CTGTATTAGA	AATAGTGGTG	TAGATGTTGT	CGTTGAAGGT	8040
	ATCGGCGACC	ATGGATTAGA	GTATATGACT	GGTGGACATG	TCATTAATTT	AGGTGATGTA	8100
	GGTAAGAACT	TCGGTCAAGG	TATGAGTGGT	GGTATTGCTT	ACGTTATCCC	GTCTGATGTA	8160
30	GAAGCTTTTG	TTGAAAATAA	TCAACTAGAT	ACGCTTTCGT	TTACAAAGAT	TAAACACCAA	8220
	GAAGAAAAAG	CATTCATTAA	GCAAATGCTG	GAAGAACATG	TGTCACACAC	GAATAGTACG	8280
	AGAGCGATTG	ATGTGTTAAA	ACATTTTGAT	CGCATTGAAG	ATGTCGTCGT	TAAAGTTATT	8340
35	CCTAAAGATT	ATCAATTAAT	GATGCAAAAA	ATTCATTTGC	ACAAATCATT	ACATGACAAT	8400
	GAAGATGAAG	CGATGTTAGC	TGCATTTTAC	GATGACAGTA	AAACAATCGA	TGCTAAACAT	8460
	AAACCAGCCG	TTGTGTATTA	AGGAAAGGGG	GAGATACGAT	GGGTGAATTT	AAAGGATTTA	8520
40	TGAAGTATGA	CAAACAGTAC	TTAGGTGAAT	TATCACTGGT	AGACCGTTTG	AAGCATCATA	8580
	AAGCATATCA	ACAACGATTT	ACTAAAGAAG	ATGCCTCTAT	CCAAGGTGCA	CGATGTATGG	8640
	ATTGTGGAAC	GCCGTTTGTG	CAAACCGGAC	AACAGTATGG	TAGGGAAACA	ATAGGTTGTC	8700
45	CAATTGGAAG	CTACATTCCCT	GAATGGAACG	ACTTAGTGTA	TCATCAAGAT	TTTAAAACTG	8760
	CTTATGAACG	CTTAAGCGAA	ACAAATAACT	TTCTGACTT	TACAGGGCGT	GTATGTCCTG	8820
50	CACCATGCGA	AAGTGCTTGT	GTGATGAAGA	TTAATAGAGA	ATCGATTGCG	ATTAAAGGTA	8880
	TTGAACGCAC	AATTATTGAT	GAAGCTTTTG	AAAATGGTTG	GGTAGCGCCG	AAAGTTCCGA	8940

55

	CTGAAGAACT TAATCTACTA GGATATCAAG TAACTATTTA TGAACGTGCT AGAGAATCAG	9060
	GCGGTTTTATT AATGTATGGT ATTCCGAATA TGAAACTTGA TAAAGATGTG GTTCGACGTC	9120
5	GTATTAAGTT AATGGAAGAA GCGGGCATT A CTTTCATTAA TGGTGTGAA GTCGGTGTG	9180
	ATATTGATAA AGCAACGTTA GAATCTGAGT ATGATGCCAT TATATTATGT ACTGGTGCAC	9240
	AAAAAGGTAG AGATTTACCT TTAGAAGGAC GCATGGGTGA TGGTATACAT TTCGCTATGG	9300
10	ATTATTTAAC TGAACAAACG CAGTTGTAA ATGGAGAAAT TGATGATATA ACAATAACTG	9360
	CAAAAGATAA GAATGTCATT ATCATTGGTG CTGGTGATAC AGGGGCAGAC TGTGTAGCGA	9420
	CAGCATTAAAG AGAAAATTGT AAATCGATTG TTCAATTTAA TAAATATACG AAATTGCCAG	9480
15	AAGCAATTAC ATTTACAGAA AATGCATCAT GGCCTTTAGC AATGCCGGTG TTTAAAATGG	9540
	ACTATGCGCA CCAAGAGTAC GAAGCTAAGT TTGGTAAGGA ACCACGTGCA TATGGTGTTC	9600
	AAACAATGCG TTACGATGTT GACGATAAAG GACACATACG TGGTTTGAT ACTCAAATTT	9660
20	TAGAGCAAGG CGAAAATGGT ATGGTCATGA AAGAAGGACC TGAAAGATTT TGGCCTGCTG	9720
	ACCTTGATTT ATTATCAATC GGCTTCGAAG GTACAGAACC AACAGTACCG AATGCTTTTA	9780
	ACATTAAAAC GGATAGAAAT CGAATCGTGG CGGATGATAC AACTATCAA ACTAATAATG	9840
25	AAAAGGTATT TGCTGCTGGA GATGCTAGAC GTGGTCAAAG TTTAGTTGTA TGGGCAATTA	9900
	AAGAAGGTAG AGGCGTAGCG AAAGCAGTAG ATCAGTATTT AGCTAGTAAA GTTTGTGTAT	9960
30	AATCTTTGTA TGGAAATGGT GGTACGTTG ACGTTGTGAC ATGCTGAATC GAGTTTGAAA	10020
	AAATCTAGTA TCTATCAACG TCACATGCCA TCTTTGTAAC CTAAAAACAA AGGTTTGTA	10080
	GACAACAAAT AGATTAATTA TAAGTAGTGA TTTTTACAT TCGTTTATAG GTCAACTGTA	10140
35	GTGGAAGACA ATGATTTGTG GTAATCATGT AATGCTTAAA AACAAATATTG ACTTTTACAG	10200
	AACGTTTATA TATGATAAAT ATTGTGTTTA GGAGGAATAC CCAAGTCCGG CTGAAGGGAT	10260
	CGGTCTTGAA AACCGACAGG GGCTTAACGG CTCGCGGGG TTCGAATCCC TCTCCTCCG	10320
40	CCATCAATAT TTATATTAAA TTCTATATAT AATGAAGTA AGTGCTCAA TTTTGAGTAT	10380
	TTACCTTTTT TATTTGTCTT TGAATGGCTC GTAATTTTGT ATAATAGAAA TGATAAGGCA	10440
	TTGAGATTGG AAGGGCATTT GGCTTGTCGA ATATACATAG CTAAATGTCT TTTTGTGTTT	10500
45	GTGAAATATG ATGGATGGCT TGTGTGACA AGTTTGCTAT TTATAGATAT GCATTTTCA	10560
	ATTTAGGAGT TGGCCATGCA TCTACACTTT ATAATGGTGA GAGCGTGGTG AGGTATTGTT	10620
	AATAACGCAA TTGTAGCGAG GAGTTATTGC TACATATGTC GTTATGGCTC ATTGATTTTC	10680
50	TGAAATGGCT ACCCCAGATA ATTGTGACAA AATAAAAATA TTTTGTGAA AGCCTTTACA	10740

55

EP 0 786 519 A2

	TAAAAAGAGA AGATGTAAAA GCCATCGTAA CCGCTATTGG GGGAAAAGAA AATCTTGAAG	10860
	CTGCAACGCA TTGTGTAACA CGATTACGTT TAGTGCTGAA GGATGAAAGT AAAGTTGATA	10920
5	AAGACGCATT AAGTAATAAC GCGTTGGTCA AGGGGCAGTT TAAAGCAGAC CATCAATATC	10980
	AAATTGTCAT TGGTCCAGGA ACAGTCGATG AAGTGTATAA GCAGTTTATT GATGAAACAG	11040
	GTGCTCAAGA AGCTTCGAAA GATGAAGCGA AACAAGCAGC TGCACAAAAA GGAATCCAG	11100
10	TACAACGTTT GATCAAATTG TGGGGGATA TTTTATACC AATATTACCT GCGATTGTGA	11160
	CAGCTGGTTT GTTAATGGGA ATCAATAATT TACTTACAAT GAAAGGTTTA TTTGGTCCAA	11220
	AAGCACTTAT TGAGATGTAT CCACAAATTG CTGATATTTC AAACATCATT AATGTGATTG	11280
15	CGAGTACGGC ATTTATTTTC TTACCAGCAT TAATTGGTTG GAGTAGTATG CGTGTATTTG	11340
	GTGGTAGTCC GATTCTAGGC ATAGTCTTAG GTTTGATTTT AATGCATCCG CAATTAGTAT	11400
	CTCAGTATGA TTTGGCAAAA GGAATATTTC CGACGTGGAA CTTATTGGC TTAGAGATTA	11460
20	AGCAGTTGAA TTACCAAGGT CAAGTGTTGC CAGTTTAAT TGCAGCTTAC GTTCTAGCTA	11520
	AAATTGAAAA AGGATTAAAT AAAGTCGTTT ACGATTGAT AAAAATGTTG GTCGTTGGAC	11580
25	CCGTAGCGCT TTTAGTTACT GGATTTTTAG CATTTATTAT CATTGGACCA GTTGCCTTAT	11640
	TGATTGGTAC AGGTATTACA TCTGGTGTTA CATTTATATT CCAACATGCA GGATGGCTTG	11700
	GCGGAGCAAT ATATGGATTG TTATATGCAC CACTTGTAAT TACAGGACTA CACCATATGT	11760
30	TTTTAGCAGT AGATTTCCAA TTGATGGGTA GCAGCTTAGG CGGTACGTAT TTATGGCCAA	11820
	TTGTTGCGAT TTCCAATATT TGTCAGGGCT CTGCAGCATT TGGAGCATGG TTTGTCTATA	11880
	AACGTCGTAA AATGGTTAAA GAAGAAGGCT TGGCATTAAAC ATCTTGTATT TCTGGTATGT	11940
35	TAGGTGTTAC TGAACCAGCC ATGTTCGGTG TGAACCTACC TCTGAAATAT CCATTTATCG	12000
	CTGCGATATC AACGTCCTGT GTATTGGGGG CAATCGTTGG TATGAATAAC GTACTTGGAA	12060
	AAGTTGGTGT TGGTGGCGTG CCAGCATTCA TTTCAATTCA AAAAGAATTT TGGCCAGTAT	12120
40	ATCTTATTGT GACAGCTATT GCTATTGTTG TACCATGTAT ACTAACAATT GTGATGTCTC	12180
	ATTTTAGTAA ACAAAAAGCG AAAGAAATTG TTGAAGATTA ATAAAATAAA AAAGGGGCGT	12240
	TCGTTATTTG GACGTCCTTT ATTACGTTAT AAGGTGGTAA TTGTGTGTCG AAAGAAATAG	12300
45	ATTGGAGAAA ATCCGTTGTA TATCAAATTT ATCCTAAGTC GTTTAATGAT ACGACGGGGA	12360
	ATGGTATAGG AGATATCAAT GGAATTATAG AAAAATTGGA TTATATCAAG TTATTGGGTG	12420
	TTGATTATAT TTGGTTAACA CCACTGTATG AATCACCAGT GAATGATAAT GGCTATGATA	12480
50	TCAGCAATTA TTTAGAAATC aATGAAGACT TTGGAACGAT GGATGATTTT GaAAAGTTAA	12540

55

EP 0 786 519 A2

	CGACGGAGCA	TGaATGGTTT	AAAGAAGCCC	GTAAATCTAA	AGATAACCCy	TATAGAGATT	12660
	ATTACTTTTT	CAGATCATCT	GAAGACGGGC	CGCCAACAAA	TTGGCATTCT	AAATTCGGTG	12720
5	GTAATGCATG	GAAGTATGAT	TCTGAGACAG	ATGAATATTA	TTTACATTTA	TTTGATGTCA	12780
	GTCAAGCTGA	TTTAAATTGG	GATAATCCGG	AAGTACGTCA	ATCGTTATAT	CGCATAGTCA	12840
	ATCATTGGAT	AGACTTCGGC	GTTGATGGTT	TTCGATTTGA	TGTCATTAAC	TTAATTTCTA	12900
10	AAGGTGAATT	TAAGGACTCT	GACAAAATAG	GTAAAGAATT	TTATACGGAT	GGTCCTAGAG	12960
	TGCATGAGTT	TCTGCATGAA	TTAAATCGTC	AAACGTTTGG	TAACACTGAC	ATGATGACTA	13020
	TAGGAGAAAT	GTCTTCGACG	ACGATTGAAA	ATTGTATTAA	GTATACACAA	CCAGAACGCC	13080
15	AAGAATTGAA	TAGTGTTTTT	AATTTTCATC	ATCTAAAGGT	TGATTATGTT	GATGGTGAAA	13140
	AGTGGACAAA	TGCGAgcTTG	nATTTTCATA	AGTTAAAGGA	AATTCTGATG	CAATGGCAAC	13200
	GAGGTATTTA	TGACGGTGGC	GGATGGAACG	CGATTTTCTG	GTGTAATCAT	GATCAGCCAC	13260
20	GGGTAGTGTC	TAGATTTGGT	GATGATACGT	CGGAAGAGAT	GAGGATACAA	AGTGCTAAAA	13320
	TGTTAGCTAT	CGCACTGCAT	ATGTTGCAAG	GGACGCCATA	TATTTACCAA	GGTGAAGAAA	13380
	TTGGTATGAC	GGACCCACAT	TTTACATCAA	TAGCACAATA	TCGTGATGTT	GAATCGATTA	13440
25	ATGCCTACCA	TCAGTTGTTA	AGTGAAGGCG	ATGCTGAAGC	GGATGTGTTA	GCGATTTTAG	13500
	GACAGAAGTC	ACGAGACAAT	TCGAGAACGC	CTATGCAATG	GAGTGATGAT	GTTAATGCTG	13560
	GATTTACAGC	TGGTAAnCCT	TGGATTGATA	TTTCGGAAAA	TTATCATCAG	GTCAACGTTA	13620
	GACAAGCACT	TCAGAATAAA	GAGTCTATTT	TCTATACGTA	TCAAAAATTA	ATACAATTAA	13680
	GACATACGCA	TGATATTATT	ACGTATGGAG	ACATTGTGCC	ACGTTTTATG	GATCATGATC	13740
35	ATTTATTTGT	TTATGAACGT	CATTATAAGA	ATCAACAATG	GCTAGTAATT	GCGAATTTC	13800
	CAGCAtCGGC	TGTTGATTTG	CCAGAAGGAT	TGGCTAGAGA	AGGTTGTGTT	GTGATTCAAA	13860
	CAGGCACAGT	GGAAAATAAT	ACGATAAGCG	GGTTTGGTGC	AATTGTAATC	GAAACAAACG	13920
40	CGTAAAATAA	ATTGAGTGGA	TGCGTTTATA	TGGCGAAACA	AAAAAAGTTT	ATGAAGATTT	13980
	ATGAGGCGTT	GAAAGAAGAT	ATATTAAACG	GGCAGATTCA	ATATGGTGAA	CAAATTCGGT	14040
	CTGAACATGA	TTTGGTGCAA	TTGTACCACT	CATCTCGAGA	GACCGTGCGT	AAGGCATTAG	14100
45	ATTTGTTGGC	ATTAGACGGC	ATGATTCAAA	AGATTCATGG	TAAAGGGTCA	CTTGTCAATT	14160
	ATCAGGAGGT	TACAGAGTTT	CCATTTTCTG	AACTTGTTAG	TTTTAAAGAA	ATGCAAGAAG	14220
	AAATGGGCGT	CGCATATTTA	ACTGAAGTTG	TTGTGAATGA	GGTTGTTGAA	GCGCATGAAG	14280
50	TTCCAGAAGT	TCAACATGCT	TTAAACATCA	ATTCTAGTGA	ATCACTCATT	CATATTGTTA	14340

55

EP 0 786 519 A2

	TTGTTTCAGA TATAGGTAAT GATGTTGCGA GTGATTCTAT TTATGATTAT TTGGAAAAGG	14460
	TATTAAATCT TAATATTAGT TATTCAAGTA AGTCTATTAC TTTTGAACCG TTTGATGAAC	14520
5	AAGCATATCA ATTGTTTGGT GATGTATCGG TGGCTTATTC AGCAACAGTT CGAAGTATTG	14580
	TGTATTTAGA AAATACAATG CCGTTTCAAT ATAATATTTC AAAACATCTT GCAAATGAAT	14640
	TTAAATTTAA TGACTTCTCA AGACGTCGTA TAAAGTAAAC AATGATATAA ATGATTTATA	14700
10	CTTGCAATTA ACTATTAAAA TATAGTAATA TATATCTTGC CGTGCTAGGT GGGGAGGTAG	14760
	CGGTTCCCTG TACTCGAAAT CCGCTTTATG CGAGGCTTAA TTCCTTTGTT GAGGCCGTAT	14820
	TTTTGCGAAG TCTGCCCAAA GCACGTAGTG TTTGAAGATT TCGGTCCTAT GCAATATGAA	14880
15	CCCATGAACC ATGTCAGGTC CTGACGGAAG CAGCATTAAAG TGGATCATCA TATGTGCCGT	14940
	AGGgTAGCCG AGATTTAGCT AACGACTTTG GTTACGTTG TGAATTACGT TCGATGCTTA	15000
	GGTGACGGT TTTTTATTTT TTAAATATTA AACCgATTAT TAAGAGTTGA AAATATATAA	15060
20	TTATAGAAGC TACTTTCTTG AAGACAATTC AGCGTATTAT ACGTGAACA TGTGTGTTGG	15120
	AAGTAGCTTT TTTATATGTG AAGTTTGATT CAAGTGAACG CGATGTGCAG TTTGAATGAT	15180
	TTTGTGTCA ATGAAAAGTA AGAAGTTATA ATTTGATGAT AAAGAAATGA TGGTGAAATG	15240
25	AGGGGGAGTA TCTTACAATA GAATTATTAA TGAGATACGT TATGATTATT GACAATCAAA	15300
	TGCCTACGGA GGACATATGC AAATATATTT AAGTACTTTA ACAGAGTTAG ATTATGATAA	15360
	ATCTTTAAAT AGTATTGAAG AAAGTTTGA TGATAATCCT GAAACGAGTT GGCAAGCACG	15420
30	TGCGAAAGTA AAACATTTAA GAAAATCTCC TTGCTATAAT TTTGAATTAG AAGTAATAGC	15480
	GAAAAATGAA AATAACGATG TCGTTGGACA CGTTTTATTA ATTGAAGTAG AAATTAATAG	15540
35	TGATGATAAG ACGTATTATG GTTTGGCGAT TGCCTCTTTA TCAGTTCATC CTGAATTACG	15600
	TGGAcAAAAA TTAGGTCGTG GCTTGGTTCA AGCAGTAGAA GAGCGTGCCA AAGCACAAGA	15660
	GTATAGTACG GTTGTGTAG ACCATTGTTT TGACTACTTT GAAAAGTTGG GTTATCAAAA	15720
40	TGCTGCTGAG CATGACATTA AATTAGAATC TGGTGATGCA CCGTTACTTG TAAAATATTT	15780
	ATGGGATAAT TTGACGGATG CACCACACGG AATCGTAAAA TTTCCAGAAC ATTTTTATTA	15840
	ATTGTTCAAT TAAGAAGTAA AGGTATTATC ATGCTATAAT GAGAGGTAAT TGTATTGGA	15900
45	GGTGCTAACT TGAATTATCA AGCCTTATAT CGTATGTACA GACCCCAAAG TTTCGAGGAT	15960
	GTCGTCGGAC AAGAACATGT CACGAAGACA TTGCGCAATG CGATTTGAA AGAAAAACAG	16020
	TCGCATGCTT ATATTTTTAG TGGTCCGAGA GGTACGGGGA AAACGAGTAT TGCCAAAGTG	16080
50	TTTGcTAAAG CAATCAACTG TCTAAATAGC ACTGATGGAG AACCTTGTA TGAATGTCAT	16140

55

	AATAATGGCG TTGATGAAAT AAGAAATATT AGAGACAAAG TTAAATATGC ACCAAGTGAA	16260
	TCGAAATATA AAGTTTATAT TATAGATGAG GTGCACATGC TAACAACAGG TGCTTTTAAAT	16320
5	GCCCTTTTAA AGACGTTAGA AGAACCTCCA GCACACGCTA TTTTATATT GGCAACGACA	16380
	GAACCACATA AAATCCCTCC AACAAATCATT TCTAGGGCAC AACGTTTTGA TTTTAAAGCA	16440
	ATTAGCCTAG ATCAAATTGT TGAACGTTTA AAATTTGTAG CAGATGCACA ACAAATTGAA	16500
10	TGTGAAGATG AAGCCTTGGC ATTTATcgCT AAAGCGTCTG AAGGGGGTAT GCGTGATGCA	16560
	TTAAGTATTA TGGATCAGGC TATTGCATTT GGTGATGGTA CGTTAACATT GCAAGATGCG	16620
	TTGAATGTCA CAGGTAGCGT ACATGATGAA GCGTTGGATC ACTTGTTTGA TGATATTGTA	16680
15	CAAGGTGACG TACAAGCATC TTTTAAAAAA TACCATCAGT TTATAACAGA AGGTAAAGAA	16740
	GTGAATCGCC TAATAAATGa TATGATTTAT TTTGTcGAG ATACGATTAT GAATAAAACA	16800
	TCTGAGAAAG ATACTGAGTA TCGAGCACTG ATGAACTTAG AATTAGATAT GTTATATCAA	16860
20	ATGATTGATC TTATTAATGA TACATTAGTG TCGATTCTGT TTAGTGTGAA TCAAAACGTT	16920
	CATTTTGAAG TGTGTTAGT AAAATTAGCT GAGCAGATTA AGGGTCAACC ACAAGTGATT	16980
25	GCGAATGTAG CTGAACCAGC ACAAATTGCT TCATCGCCAA ACACAGATGT ATTGTTGCAA	17040
	CGTATGGAAC AGTTAGAGCA AGAACTAAAA ACACTAAAAG CACAAGGAGT GAGTGTCGCT	17100
	CCTGTTCAAA AATCTTCGAA AAAGCCTGCG AGAGGCATAC AAAAATCTAA AAATGCATTT	17160
30	TCAATGCAAC AAATTGCAAA AGTGCTAGAT AAAGCGAATA AGGCAGATAT CAAATTGTTG	17220
	AAAGATCATT GGCAAGAAGT GATTGATCAT GCCAAAAATA ATGATAAAAA ATCACTCGTT	17280
	AGTTTATTGC AAAATTCGGA ACCTGTGGCG GCAAGTGAAG ATCACGTACT TGTGAAATTT	17340
35	GAGGAAGAGA TCCATTGTGA AATCGTCAAT AAAGACGACG AGAAACGTAG TAGTATAGAA	17400
	AGTGTGTAT GTAATATCGT TAATAAAAAAC GTTAAAGTTG TTGGTGTACC ATCAGATCAA	17460
	TGGCAAAGAG TTCGAACGGA ATATTTACAA AATCGTAAAA ACGAAGGCGA TGATATGCCA	17520
40	AAGCAACAAG CACAACAAAC AGATATTGCT CAAAAAGCAA AAGATCTTTT CGGTGAAGAA	17580
	ACTGTACATG TGATAGATGA AGAGTGATAC ATGACAAGCG ATATAATCGT ATGTATAATG	17640
	AAAGAAACAT CATTTTATTG ATAAATATTT ATTGATTTTC AAGGAGGAAA TGGAATATGC	17700
45	GCGGTGGCGG AAACATGCAA CAAATGATGA AACAAATGCA AAAAATGCAA AAGAAAATGG	17760
	CTCAAGAACA AGAAAACTT AAAGAAGAGC GTATTGTAGG AACAGCTGGC GGTGGCATGG	17820
	TTGCAGTTAC TGTAACGGT CATAAAGAAG TTGTCGACGT TGAAATCAAA GAAGAAGCTG	17880
50	TAGACCCAGA CGATATTGAA ATGCTACAAG ACTTAGTGTT AGCAGCTACT AATGAAGCGA	17940

55

EP 0 786 519 A2

	TCCCTGGaAT GTGATCATAG ATGCATTATC CAGAACCTAT ATCAAACTT ATTGATAGCT	18060
	TTATGAAATT GCCAGGCATT GGTCCAAAGA CAGCCCAACG TCTGGCTTTT CATACCTTAG	18120
5	ATATGAAAGA AGACGATGTT GTTCAGTTTG CCAAAGCATT AGTAGATGTT AAGAGAGAAT	18180
	TAACATATTG TAGCGTATGT GGTACACATTA CTGAAAATGA TCCATGTTAT ATTTGTGAAG	18240
	ATAAGCAAAG AGATCGTTCA GTTATTTGTG TTGTGGAAGA TGACAAAGAT GTCATAGCTA	18300
10	TGGAAAAAAT GAGAGAATAC AAAGGTTTAT ATCACGTTTT ACATGGGTCT ATTTTCGCCTA	18360
	TGGATGGCAT TGGACCAGAA GATATTAATA TTCCTTCATT GATTGAACGC TTGAAAAACG	18420
	ATGAAGTTAG CGAATTAATC TTAGCTATGA ACCCGAACTT AGAGGGGGAA TCTACAGCCA	18480
15	TGTATATTTT TAGATTAGTT AAGCCTATAG GTATCAAAGT GACGAGATTA GCACAAGGGT	18540
	TATCGGTAGG TGGCGATTTA GAGTATGCTG ACGAAGTAAC ATTATCTAAA GCAATCGCAG	18600
	GTAGAACAGA AATGTAATkT CTTCTATTAA ACATTTTTGA TTTTAATACT ATAGTAAGAA	18660
20	AAGTCACAGT GTAATCATTG TGGCTTTTTT TATGGTGTGG TGTGATGTAC TACTTTATTT	18720
	GCGGTGTGGC GGTGGTATGG TTTACCTAGT TTTACTGAGG GATGGGTAAT CTTTAGGAAG	18780
	CAAGCCGTTG GTTGTGATTT GTTACTTCTA ATAGTAATGA TGTGAATTGG ATTATCGAAT	18840
25	TAGATCTATG GTTATGGTGT GTTGGTGCTA TTAATTGAT AAATGCGGTT AATGACTATG	18900
	CAAATGAAAT TCTTTTGTA TGAATGAT AGATGCTGGC TTAGTAAGTT GTACTTCTTT	18960
30	GGTCTAAAGC TTATTAAATC AGCCTGTATA GCGGTGTTTT GAGAGATTAT TTAAACTTG	19020
	TAAATTTATT TTTAATTCTT GGTAAAAAAA TAACGTCTCG TTTTGCCTTT TTTTGTATTG	19080
	ATATGGTTAG AGAAAAATCT GTTCTTGTT CTAAAAACG TACTATTTAT AAGTGGGGAT	19140
35	TTTTTAAGTT CGATTTTATG GATAAGGGCG TTCAGTACAG ATGACAAAGG TGTAAATTTT	19200
	ACTGTGTGTA AGCAGTTTGA AAGCCTGTAT AGTATTTATT TGTGAGGCA AACAAAACAA	19260
	CTCAACTTAA GAAATAACTT GAATTACTAA CGAAAATTAA TTTTAAAAAG TTATTGACTT	19320
40	AAATGTTAAT AAAATGTATA ATTAATTCTT GTCGGTAAGA AAAATGAACA TTGAAACTG	19380
	AATGACAATA TGTCAACGTT AATTCCAAAA AACGTAAC TAAGTTACAA ACATTATTTA	19440
	GTATTTATGA GCTAATCAAA CATCATAATT TTTATGGAGA GTTTGATCCT GGCTCAGGAT	19500
45	GAACGCTGGC GCGTGCCTA ATACATGCAA GTCGAGCGAA CGGACGAGAA GCTTGCTTCT	19560
	CTGATGTTAG CGGCGGACGG GTGAGTAACA CGTGGATAAC CTACCTATAA GACTGGGATA	19620
	ACTTCGGGAA ACCGkAGCTA ATACCGGATA ATATTTTGAA CCGCATGGTT CAAAAGTGAA	19680
50	AGACGGTCTT GCTGTCACTT ATAGATGGAT CCGCGCTGCA TTAGCTAGTT GGTAAGGTAA	19740

55

EP 0 786 519 A2

	GAGACACGGT	CCAGACTCCT	ACGGGAGGCA	GCAGTAGGGA	ATCTTCCGCA	ATGGGCGAAA	19860
	gCtGaCGGAG	CAACGCCGCG	TGAGTGATGA	AGGTCTTCGG	ATCGTAAAAC	TCTGTTATTA	19920
5	GGGAAGAACA	TATGTGTAAG	TAAGTGTGCA	CATCTTGACG	GTACCTAATC	AGAAAGCCAC	19980
	GGCTAACTAC	GTGCCAGCAG	CCGCGGTAAT	ACGTAGGTGG	CAAGCGTTAT	CCGGAATTAT	20040
	TGGGCGTAAA	GCGCGCGTAG	GCGGTTTTTT	AAGTCTGATG	TGAAAGCCCA	CGGCTCAACC	20100
10	GTGGAGGGTC	ATTGGAAACT	GGAAAACCTG	AGTGCAGAAG	AGGAAAGTGG	AATTCCATGT	20160
	GTAGCGGTGA	AATGCGCAGA	GATATGGAGG	AACACCAGTG	GCGAAGGCGA	CTTTCTGGTC	20220
	TGTAAGTGAC	GCTGATGTGC	GAAAGCGTGG	GGATCAAACA	GGATTAGATA	CCCTGGTAGT	20280
15	CCACGCCGTA	AACGATGAGT	GCTAAGTGTT	AGGGGGTTTC	CGCCCTTAG	TGCTGCAGCT	20340
	AACGCATTAA	GCACTCCGCC	TGGGGAGTAC	GACCGCAAGt	TGAAACTCAA	AGGAATTGAC	20400
	GGGACCCGCG	ACAAGCGGTG	GAGCATGTGG	TTTAATTCGA	AGCAACGCGA	AGAACCCTAC	20460
20	CAAATCTTGA	CATCCTTTGA	CAACTCTAGA	GATAGAGCCT	TCCCCTTCGG	GGGACAAAGT	20520
	GACAGGTGGT	GCATGGTTGT	CGTCAGCTCG	TGTCGTGAGA	TGTTGGGTTA	AGTCCCGCAA	20580
25	CGAGCGCAAC	CCTTAAGCTT	AGTTGCCATC	ATTAAGTTGG	GCACTCTAAG	TTGACTGCCG	20640
	GTGACAAACC	GGAGGAAGGT	GGGGATGACG	TCAAATCATC	ATGCCCTTA	TGATTTGGGC	20700
	TACACACGTG	CTACAATGGA	CAATACAAAG	GGCAGCGAAA	CCGCGAGGTC	AAGCAAATCC	20760
30	CATAAAGTTG	TTCTCAGTTC	GGATTGTAGT	CTGCAACTCG	ACTACATGAA	GCTGGAATCG	20820
	CTAGTAATCG	TAGATCAGCA	TGCTACGGTG	AATACGTTCC	CGGGTCTTGT	ACACACCGCC	20880
	CGTCACACCA	CGAGAGTTTG	TAACACCCGA	AGCCGGTGGA	GTAACCTTTT	AGGAGCTAGC	20940
35	CGTCGAAGGT	GGGACAAATG	ATTGGGGTGA	AGTCGTAACA	AGGTAGCCGT	ATCGGAAGGT	21000
	GCGGCTGGAT	CACCTCCTTT	CTAAGGATAT	ATTCGGAACA	TCTTCTTCAG	AAGATGCGGA	21060
	ATAACGTGAC	ATATTGTATT	CAGTTTTGAA	TGTTTATTTA	ACATTCAAAT	ATTTTTTGGT	21120
40	TAAAGTGATA	TTGCTTATGA	AAATAAAGCA	GTATGCGAGC	GCTTGACTAA	AAAGAAATTG	21180
	TACATTGAAA	ACTAGATAAG	TAAGTAAAT	ATAGATTTTA	CCAAGCAAAA	CCGAGTGAAT	21240
	AAAGAGTTTT	AAATAAGCTT	GAATTCATAA	GAAATAATCG	CTAGTGTTTC	AAAGAACACT	21300
45	CACAAGATTA	ATAACGCGTT	TAAATCTTTT	TATAAAAGAA	CGTAACTTCA	TGTTAACGTT	21360
	TGACTTATAA	AAATGGTGGA	AACATAGATT	AAGTTATTAA	GGGCGCACGG	TGGATGCCTT	21420
	GGCACTAGAA	GCCGATGAAG	GACGTTACTA	ACGACGATAT	GCTTTGGGGA	GCTGTAAGTA	21480
50	AGCTTTGATC	CAGAGATTTT	CGAATGGGGA	AACCCAGCAT	GAGTTATGTC	ATGTTATCGA	21540

55



EP 0 786 519 A2

	GAGGAAGAGA	AAGAAAATTC	GATTCCCTTA	GTAGCGGCGA	GCGAAACGGG	AAGAGCCCAA	21660
	ACCAACAAGC	TTGCTTGTTG	GGGTTGTAGG	ACACTCTATA	CGGAGTTACA	AAGGACGACA	21720
5	TTAGACGAAT	CATCTGGAAA	GATGAATCAA	AGAAGGTAAT	AATCCTGTAG	TCGAAAATGT	21780
	TGTCTCTCTT	GAGTGGATCC	TGAGTACGAC	GGAGCACGTG	AAATTCCGTC	GGAATCTGGG	21840
	AGGACCATCT	CCTAAGGCTA	AATACTCTCT	AGTGACCGAT	AGTGAACCAG	TACCGTGAGG	21900
10	GAAAGGTGAA	AAGCACCCCG	GAAGGGGAGT	GAAATAGAAC	CTGAAACCGT	GTGCTTACAA	21960
	GTAGTCAGAG	CCCGTTAATG	GGTGATGGCG	TGCCTTTTGT	AGAATGAACC	GGCGAGTTAC	22020
	GATTTGATGC	AAGGTTAAGC	AGTAAATGTG	GAGCCGTAGC	GAAAGCGAGT	CTGAATAGGG	22080
15	CGTTTAGTAT	TTGGTCGTAG	ACCCGAAACC	AGGTGATCTA	CCCTTGGTCA	GGTTGAAGTT	22140
	CAGGTAACAC	TGAATGGAGG	ACCGAACCGA	CTTACGTTGA	AAAGTGAGCG	GATGAACTGA	22200
	GGGTAGCGGA	GAAATTCCAA	TCGAACCTGG	AGATAGCTGG	TTCTCTCCGA	AATAGCTTTA	22260
20	GGGCTAGCCT	CAAGTGATGA	TTATTGGAGG	TAGAGCACTG	TTTGGACGAG	GGGCCCCCTCT	22320
	CGGGTTACCG	AATTCAGACA	AACTCCGAAT	GCCAATTAAT	TTAACTTGGG	AGTCAGAACA	22380
25	TGGGTGATAA	GGTCCGTGTT	CGAAAGGGAA	ACAGCCCAGA	CCACCAGCTA	AGGTCCCAAA	22440
	ATATATGTTA	AGTGGAAAAG	GATGTGGCGT	TGCCCAGACA	ACTAGGATGT	TGGCTTAGAA	22500
	GCAGCCATCA	TTTAAAGAGT	GCGTAATAGC	TCACTAGTCG	AGTGACACTG	CGCCGAAAAT	22560
30	GTACCGGGGC	TAAACATATT	ACCGAAGCTG	TGGATTGTCC	TTTGGaCAAT	GgTAGGAGAG	22620
	CGTTCTAAGG	GCGTTGAAGC	ATGATCGTAA	GGACATGTGG	AGCGCTTAGA	AGTGAGAATG	22680
	CCGGTGTGAG	TAGCGAAAAG	CGGGTGAGAA	TCCCGTCCAC	CGATTGACTA	AGGTTTCCAG	22740
35	AGGAAGGCTC	GTCCGCTCTG	GGTtagTCGG	GTCCTAAGCT	GAGGCCGACA	GcGTAGGCGA	22800
	TGGAFAACAG	GTTGATATTC	CTGTACCACC	TATAATCGTT	TTAATCGATG	GGGGGACGCA	22860
	tAGGATAGGC	GAAGcGTGcG	ATTGGATTGC	ACGTCTAAGC	AGTAAGGCTG	AGTATTAGGC	22920
40	AAATCCGGTA	CTCGTTAAGG	CTGAGCTGTG	ATGGGGAGAA	GACATTGTGT	CTTCGAGTCG	22980
	TTGATTTTAC	ACTGCCGAGA	AAAGCCTCTA	GATAGAAAAT	AGGTGCCCCG	ACCGCAAACC	23040
	GACACAGGTA	GTCAAGATGA	GAATTCTAAG	GTGAGCGAGC	GAACTCTCGT	TAAGGAACTC	23100
45	GGCAAAATGA	CCCCGTAAC	TCGGGAGAAG	GGGTGCTCTT	TAGGGTTAAC	GCCCAGAAGA	23160
	GCCGCAGTGA	ATAGGCCCAA	GCGACTGTTT	ATCAAAAACA	CAGGTCTCTG	CTAAACCGTA	23220
	AGGTGATGTA	TagGGcTGAC	GCCTGCCCCG	TGCTGGAAGG	TTAAGAGGAG	TGGTTAGcTT	23280
50	CTGCGAAGCT	ACGAATCGAA	GCCCCAGTAA	ACGGCGGCCG	TAACTATAAC	GGTCCTAAGG	23340

55

EP 0 786 519 A2

	TGTCTCAACG	AGAGACTCGG	TGAAATCATA	GTACCTGTGA	AGATGCAGGT	TACCCGCGAC	23460
	AGGACGGAAA	GACCCCGTGG	AGCTTTACTG	TAGCCTGATA	TTGAAATTCG	GCACAGCTTG	23520
5	TACAGGATAG	GTAGGAGCCT	TTGAAACGTG	AGCGCTAGCT	TACGTGGAGG	CGCTGGTGGG	23580
	ATACTACCCT	AGCTGTGTTG	GCTTTCTAAC	CCGCACCACT	TATCGTGGTG	GGAGACAGTG	23640
	TCAGGCGGGC	AGTTTGGACTG	GGGCGGTGCG	CTCCTAAAAG	GTAACGGAGG	CGCTCAAAGG	23700
10	TTCCCTCAGA	ATGTTTGAA	ATCATTCATA	GAGTGTAAG	GCATAAGGGA	GCTTGACTGC	23760
	GAGACCTACA	AGTCGAGCAG	GGTCGAAAGA	CGGACTTAGT	GATCCGGTGG	TTCCGCATGG	23820
	AAGGGCCATC	GCTCAACGGA	TAAAAGCTAC	CCCGGGGATA	ACAGGCTTAT	CTCCCCCAAG	23880
15	AGTTCACATC	GACGGGGAGG	TTTGGCACCT	CGATGTCGGC	TCATCGCATC	CTGGGGCTGT	23940
	AGTCGGTCCC	AAGGGTTGGg	CTGTTGCCCC	ATTAAAGCGG	TACGCGAGCT	GGGTTCAGAA	24000
	CGTCGTGAGA	CAGTTCGGTC	CCTATCCGTC	GTGGGCGTAG	GAAATTTGAG	AGGAGCTGTC	24060
20	CTTAGTACGA	GAGGACCGGG	ATGGACATAC	CTCTGGTGTA	CCAGTTGTCTG	TGCCAACGGC	24120
	ATAGCTGGGT	AGCTATGTGT	GGACGGGATA	AGTGCTGAAA	GCATCTAAGC	ATGAAGCCCC	24180
	CCTCAAGATG	AGATTTCCCA	ACTTCGGTTA	TAAGATCCCT	CAAAGATGAT	GAGGTTAATA	24240
25	GGTTCGAGGT	GGAAGCATGG	TGACATGTGG	AGCTGACGAA	TACTAATCGA	TCCAAGACTT	24300
	AATCAAAATA	AATGTTTTGC	GAAGCAAAAT	CACTTTTACT	TACTATCTAG	TTTTGAATGT	24360
	ATAAATTACA	TTCATATGTC	TGGTGACTAT	AGCAAGGAGG	TCACACCTGT	TCCCATGCCG	24420
	AACACAGAAG	TTAAGCTCCT	TAGCGTCGAT	GGTAGTcGAA	CTTACGTTCC	GCTAGAGTAG	24480
	AACGTTGCCA	GGCAAAAAAT	GGATGCGATG	AGCCGCATTG	AGACCGCAAG	GTCTCTTTTT	24540
35	TTTATGTCTA	AAACGTCAAA	ATAAAAAGCA	AACACAAAGA	AAAATGGCTT	GGCGAAGTGA	24600
	AAACGTTTGA	ATCTGACGAA	ACGAGAAAAG	AfCGCAACGA	GTTTAGTAGA	GCTAAATGAG	24660
	TAAGyGAGAG	CCGAAGrAGA	GGAAAGAAGC	AAGCGATTGT	CACAAGTCAA	GAAAGGTTCT	24720
40	TAGCGAsGAT	GGTAGCCAAC	TTACGTTCCG	CTAGAGTAGA	ACTGGAAATG	ATAATTTAAT	24780
	AATGTACACT	TTCGATTGTC	TAAGTATGTA	CAACTTTAAT	TTTGTGTTTA	TATAAATTTA	24840
	AAATGATATC	ATCGAAAACA	AAATATTGTA	TAAATAGAGA	AGAGCAGTAA	GACGGTATCT	24900
45	AATTGAAAAT	GATCTTACTG	CTCTTTTATA	TACTTTATTG	AAATACAAAA	AGGAAATTAA	24960
	TTATTATACA	ATAGACAAGC	TATTGCATAA	GTAACACTAA	CTTTTATCAA	AGAAGTGTTA	25020
	CTTTATAATT	AATGATTTTA	TTAGAGCGTC	TACATGCGGT	TTTAAAGCAT	CATCGTCTAT	25080
50	ACCGCCAAAG	CCTAATATAA	ATTTAGGGGT	TTTCTTATAG	TCTTGATCAT	CATCAAAATT	25140

55

EP 0 786 519 A2

	TCCATTTTTT	ACTGTAATTG	TAAAATGCAT	ACCCGTTTCA	GCACCTTGAA	TATCAAGCTG	25260
	CTCTTTGTAA	GGTTTCAATC	TTTTTAAAT	ATAGGTTACT	TTTCTACGAT	AAATTCGTCT	25320
5	CATTTTATTT	AAATGCCTTT	CAAAACCACC	GGAAGATATA	AACGTTGCAA	TAAGGTTTTG	25380
	CATATGAACA	GGTACAGTGT	TGCCTTCAAT	GTGATTTTGA	GAATGATATT	TTTTCATTAT	25440
	AGAATAGGGT	AACACCATAT	ATGCAACTCG	ACAGCTAGGA	AAAATAGACT	TGAAAATGT	25500
10	ACTGATATAA	ATCACTTTTT	CTCCTCTTGA	ATATAGACCT	TGAATGCTG	GAATGGGTTT	25560
	GCCGAAATAT	CTAAACTCGG	AATCATAATC	ATCTTCTATA	ATAAATCGTT	CTTCTTTTTC	25620
	TTGAGCCCAT	TGTATTAATT	GAGTTCGTTT	TTTTAAGTCC	ATCACATATC	CAGTTGGAAA	25680
15	TTGATGGGAA	GGCGTTATAT	ATACTATATT	TTTTTGTGAT	TTAATAACTT	CATCTACGTT	25740
	TATTCCATTA	TCTTCAACTT	CAATTTGTTT	ATATTCAACT	TGTTTTTTAT	CTAAAATATT	25800
	TTTGATTGGT	GGATAACTAG	GTTTTTCGAT	AATAAATGTT	GAAGTATAAA	GTAAATCGAC	25860
20	TAATTGATTT	ACTAATTGTT	CGGTAGATGA	GCCAATTATA	ATTTGATTAG	GATCACAAAT	25920
	TACGCCACGA	TTAGTAAATA	AATAAAATGC	CAGTTGAAAC	CGCAAATGTA	ATTCTCCTTG	25980
25	AAAATGTCCT	CTACGTAATT	GATTTAAATG	ATTTGTATCA	TAAAGATCTT	TGGAATACTT	26040
	TCTGAAAAGT	TCTATAGGGA	AATGTTTCGT	ATCTATTTCA	TCCAAATTAA	AAGCATAATC	26100
	ATAAGCTTCA	TCACTCGCTT	TTGGTTTATA	TGAATCATCA	TCAAAAAGAG	AGGGGATAGG	26160
30	TTGATTGTTT	AAAATTGTTA	AAGATTCAAT	TTCGGACACA	AAATATCCAG	AGCGAGGTCT	26220
	TGAATAAATG	TAACCTTCGT	CTAATAGAAG	TTGATATGCA	TGCTCTACGG	TTGTTTGGCT	26280
	AATAGATAAA	TGTTTGCTTA	ATTGTCTTTT	AGAATAAAAT	TTATCGCCTT	CTTTAAATTG	26340
35	ACCTTCAATT	ATTTGTTTTT	TTAATTTTTT	ATAAAGTTGA	TGGTATAAAG	TGTTTTTCAA	26400
	TTTTATAACT	GACCTCCTAA	ATTTATCTTA	TTTTGTACCT	TTTTAAATAT	CAGTTTATAC	26460
	ATTACAATGT	ATTTAATCAA	CTTGAAAAGG	GGTTTTATGT	ATAATGAGTA	AAATTATTGG	26520
40	ATCAGACAGA	GTCAAAAGAG	GTATGGCTGA	AATGCAAAAA	GGCGGCGTTA	TTATGGATGT	26580
	CGTTAATGCT	GAGCAAGCAA	GAATTGCAGA	AGAAGCTGGC	GCGGTAGCAG	TTATGGCATT	26640
	AGAACGAGTA	CCTTCTGATA	TTAGAGCTGC	TGGTGGTGT	GCACGTATGG	CAAACCCTAA	26700
45	AATTGTAGAA	GAAGTAATGA	ATGCTGTTTC	TATTCCAGTC	ATGGCTAAAG	CACGTATTGG	26760
	TCATATCACT	GAAGCAAGAG	TATTAGAGGC	GATGGGTGTT	GACTATATTG	ATGAATCAGA	26820
	AGTGTTAACA	CCAGCAGATG	AGGAATATCA	CTTAAGAAAA	GATCAATTTA	CAGTACCATT	26880
50	TGTATGTGGA	TGTCGTAATT	TAGGTGAAGm	TGCGCGTAGA	ATTGGTGAAG	GTGCTGCTAT	26940

55

	ACAAGTTAAT TCAGAAGTTA GTCGATTGAC TGTAAATGAAT GATGATGAGA TTATGACTTT	27060
	TGCGAAAGAT ATCGGTGCGC CTTATGAAAT TTTAAACAA ATTAAGACA ATGGTCGTTT	27120
5	ACCGGTAGTT AACTTTGCAG CTGGTGGCGT TGCGACTCCT CAAGATGCTG CTTTAATGAT	27180
	GGAATTAGGT GCTGACGGTG TATTCGTTGG ATCAGGTATT TTTAAATCAG AAGATCCAGA	27240
	AAAAATTGCT AAAGCAATTG TTCAAGCAAC AACACATTAC CAAGACTATG AACTAATTGG	27300
10	AAGATTAGCA AGTGAAGTTG GCACTGCTAT GAAAGGTTTA GATATCAATC AATTATCATT	27360
	AGAAGAACGT ATGCAAGAGC GTGGTTGGTA AGATATGAAA ATAGGTGTAT TAGCATTACA	27420
15	AGGTGCAGTA CGTGAACATA TTAGACATAT TGAATTAAGT GGTCAATGAAG GTATTGCAGT	27480
	TAAAAAAGTT GAACAATTAG AAGAAATCGA GGGCTTAATA TTACCTGGTG GCGAGTCTAC	27540
	AACGTTACGT CGATTAATGA ATTTATATGG ATTTAAAGAG GCTTTACAAA ATTCAACTTT	27600
20	ACCTATGTTT GGTACATGCG CAGGATTAAT AGTTCTAGCG CAAGATATAG TTGGTGAAGA	27660
	AGGATACCTT AACAAAGTTGA ATATTACTGT ACAACGAAAC TCATTCGGTA GACAAGTTGA	27720
	CAGCTTTGAA ACAGAATTAG ATATTAAAGG TATCGCTACA GATATTGAAG GTGTCTTTAT	27780
25	AAGAGCCCCA CATATTGAAA AAGTAGGTCA AGGCGTAGAT ATCCTATGTA AGGTAAATGA	27840
	GAAAATTGTA GCTGTTGAGC AAGGTAAATA TTTAGGCGTA TCATTCCATC CTGAATTAAC	27900
	AGATGACTAT AGAGTAACTG ATTACTTTAT TAATCATATT GTAAAaAAG CATAGCTTAA	27960
30	TGTATGCTAA ATCAACGAAT TATTGATATT TATAGATTTG TTGAGAAGAA AATATCTCCT	28020
	TCAAACCTAG CTTTGGAGGA GTTATTTTTT ATGTCAAAAT TAAAAATGAT AAAAAATAAA	28080
35	GCTATACATA AGAAAAAAC CCTTCAAAGA GACTGAGAAT AGTCAAAATT TTGAAGGGGT	28140
	TAATTGCATG TTGATGTATT TGTTAAATAA AGAATCcAGC GATTGCAGCT GAAATGAAAG	28200
	ATACTAGTGT tGCACCGAAT AATAATTTCA AACCAAAGCG GGCAACTGTA TCTCCTTTTT	28260
40	TGTCATTAAG TGATTTAATC GCACCTGAAA TAATACCGAT AGAGCTAAAG TTAGCAAATG	28320
	ATACTAAGAA TACAGATGTA ACACCTTTTG CGTGTTCAGA TAAATCACTA AGTTTACCAA	28380
	GTGCTTGCAT TGCTACAAAT TCGTTAGATA ATAGTTTTGT CGCCATAACT GAACCGGCTT	28440
45	GAACTGCATC TTGCCATGGC ACACCGACTA AGAATGCAAA TGGTGCAAAG ACAAACCAA	28500
	TTAATGTTTG GAAATCCCAA GAAATAGCGC CACCTGAAAC TGTACTAAAG ATATTGCTTA	28560
	CAATTCCATT TAATAGAGCG ATAATGGCAA TGTATCCGAT TAACATTGCG CCTACAATGA	28620
50	CAGCTACTTT AAATCCATCT AAAATATATT CTCCTAGCAT TTCGAAGAAT GATTGTTGTC	28680
	TTTCTTCAGT TTCTTCAACT AATAATTGT CATCTTCTTC ATTAACTTTA TAAGGGTTAA	28740
55		

	TAGGTTCAAT TAAGGTAAAG TATGCACCGA TAATTGAAGC AGAAACAGTC GACATTGCTG	28860
	AAGCTGTTAA TGTGTATAAA CGTTGCTTAG GTATGTATGG TAATTGTTTT TTAATTGAAA	28920
5	TAAATACTTC AGATTGTCCC AAAATTGCTG CAGCAACTGC ATTGTATGAT TCTAAACGTC	28980
	CCATACCATT AATTTTAGAA ATTAAGAATC CTAAAACATT AATGATTAAA GGTAATCT	29040
	TTGTGTATTG AAGGATACCG ATAATCGCTG AAATAAATAC GATAGGTAAT AATACACTGA	29100
10	AGAAGAATGG TGGTTGCTTA GGATCGATAT ATTGAATACC ACCGAATACA AAGTTAACAC	29160
	CATCTGCTGC TTTTAATAAT AAGTAGTTAA AACCCTTTGA AATACCACCA ATAACCTTGA	29220
	TTCCCATTGT AGTTTTAAGC AAGATAAATG CAAAGATAAG CTGAATTGCA AGTAAAATTC	29280
15	CTACATATTT CCAGCGAATA TTTTCTCTGT CTGAGCTAAA TAGAAACGCA AGTGCTAAAA	29340
	AGAAGATAAT TCCGATAATC CCAATTAGAA TATGCATATA TTTCTCATT CTTTAGTTTT	29400
20	TTCTACaATc TATCATACAA TAAATGGAA GGGCTAACAT CATAAATTTT TGAAAATATA	29460
	AAAACAAATT AATTGAAAAA GGTCAAAATA GGTCATATAA TATAGTCAAA GAAGGTCAAA	29520
	AAGGGGTGAT ATACATGCAC AATATGTCTG ACATCATAGA ACAATAaTCA AACGTTTATT	29580
25	TGAAGAGTCG AATGAAGATG TCGTTGAAAT TCAGAGAGCG AATATCGCAC AGCGTTTTGA	29640
	TTGCGTACCA TCACAATTAA ATTATGTAAT CAAAACACGA TTCACTAATG AACATGGTTA	29700
	TGAAATCGAA AGTAAACGTG GTGGTGGTGG TTACATCCGA ATCACTAAAA TTGAAAATAA	29760
30	AGATGCAACA GGTTATATTA ATCATTTGCT TCAGCTGATT GGACCTTCTA TTTCTCAACA	29820
	ACAAGCTTAT TATATTATTG ATGGGCTTTT AGATAAAATG TTAATAAATG AACGTGAAGC	29880
	TAAATGATT CAAGCAGTTA TTGATAGAGA AACGCTATCA ATGGATATGG TTTCTAGAGA	29940
35	TATTATTAGA GCAAATATTT TAAACGTTT GTTACCAGTT ATAAATTATT ACTAAATGAA	30000
	ATGAGGTGTT GAAGTGCTTT GTGAAAATTG TCAACTTAAT GAAGCGGAAT TAAAAGTTAA	30060
40	AGTTACAAGT AAAAATAAAA CAGAAGAAAA AATGGTGTGT CAACTTGTG CTGAGGGGCA	30120
	CCATCCGTGG AATCAAGCTA ATGAACAACC TGAaTATCAA GAACATCAAG ATAATTTCGA	30180
	AGAAGCATT GTTGTAAAGC AAATTTTACA ACATTTAGCT ACGAAACATG GAATTAATTT	30240
45	TCAAGA	30246

(2) INFORMATION FOR SEQ ID NO: 57:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14333 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

	TATTCCCCCA TCGGTTTATT AAATCGTCCA TTCAATACT GTTTTCCCCC AAGATGTCGA	60
5	TAAATCCATT TCAAACGCTT GGACGATATC TTGCATCGTA CATACATTAA TTTCATGTCC	120
	TTTTAATAAT GCTAACTTTT CAACTATGTC TGGGTACTTA CGATATAAAT CAACAACCTG	180
	CTCAAAATCT TTAGAGCCGC TTCGACTACT ACCAATCAAC GTTAATCCTT TTTCAAGTAC	240
10	TAATCGTGTA TTCACTTCCA CGGGTAATTC ACTTACGCCT AACAAAGCAA TACTGCCTTC	300
	TGGTGAAATA TGTTCAACTA TTTGTTGAAG TGCAACTTGA CTTCTTTTAC CTCCAACACA	360
15	TTCAAATGCA TGATCAATTT TAAGATCATC TGGTATTTGA TTTACTGTAA AGATGTCATC	420
	TACAAATGAA AAATGACTTA ATTTATAGTC TGTCTTACCA AATACATAAG TTTTAGCTTC	480
	TGGGTACAAC TTACGTAGCA AAATAGCAGT AATATAACCT AAGTTACCAT CACCCCAAAT	540
20	ACCAAAGCTG GTTTTCAAAG GTATAGATTT ACGTTCAAAT CGTTGTATAG CATGATAACT	600
	TACTGACACT AACTCTGTGT ATGAAATCGT ACTCAAATCA ATGTCATTAG GCAGCGGAAC	660
	GATACGATCA TGTGCCATCA CAACGTAGTC TTGCATAAAA CCATCATAAC CACTAGATCT	720
25	AAAATAACTA GAGGCTAAGT AATTCTCCGC AATAATATGA TGTGCTCTG TAGGTGTATT	780
	CGGTACCATT ACTACTTTCG TACCTTTTTT AAATACCCCT TTACTATCAA ATACAACCTC	840
	ACCAACAGCT TCATGAACTA ATGACATTGG TAATTTTTTG CGTAGTACAT TTTCATCTCT	900
30	TCGACCTGTG TAATACCTTT GATCAGCTGC ACAAATAGAC AAGTATAAAG GTCTTACGAT	960
	GACATGATTA CCATAAATAT CAACATTATT ATATGTGACG TCGAACTGTC TCGGTGCAAC	1020
35	GAGTTGATAT ACTTGATTAA TCATCGGCAA TATCACCTTG AATAATGGCA TTTGCTACTT	1080
	TTAAATCATA CGGTGTTGTC ACTTTAATGT TGTATAGTTC TCCaCGTACC AATTTAACTG	1140
	CATGTCCAGA TTCGACAAATG ATTTTACATG CATCTGATAA GATTTCTTTT TGTTCACTAC	1200
40	TTAAGGCGCG ATAACATCTT TGTAATAATT TAATATTAAA TGATTGTGGT GTTTGGCCTT	1260
	GATACATTTC ATTCCTTACA GGGATACTGT GTATGTTCTG TTTATCTTTA GACATTACAA	1320
	TCGTATCAAT TGCTTCAATG ACTGTATCTA CTGCACCATA TTTTGCTGCT ACTTCAATGT	1380
45	TCTCTTTAAT AATACGTTGA GTTAAAAATG GTCTTACGGC ATCATGAGTT ACAATCACAT	1440
	CATCATTATT AATTCCATTT ACATTGCGAA TATGGTCGAT AATGTTTATA ATTGTTTCAT	1500
	TTGATCCGT ACCACCTGCA ACTACTTTGA CACGTTGATC TGTAATGTTA TATTTTTTTA	1560
50	AAATATCCTG TGTATGGGAA ATCCACTGTG CTGGCGTTGC GATAATAATC TCATTAAATT	1620
	CACTCACTAA AATGAACTTC TCAATTGTAT GGATTAAAAT CGGTTTATTA TCAATATCTA	1680

55

	CTGCATAAAT CATGTTGTCC TCCATTCTGT CATTACATCA TTTCCATTTA TACATTACTG	1800
	ACCTATGCCC GCACATAAGC CTAACCTATT GCTCACTTGC CTCTTTTATT AATCCAAAGA	1860
5	TAGTTGTCAC AATAGTGTGA TAATTTTTTA TAAAAATGTA TTTTGTAAAC TGACCATTCT	1920
	AAGTTGTTTT GCCATGCAGT TAATCATTAA CTCTGACGAT ATTAAATTGT TAAAGGTATT	1980
10	AATGTTTACT CTTTTTCAAA TTCATTATTA CTGCCATCAT TTTACCATAT ATTATAATAA	2040
	ATTTATCTTA TTAAGTGGCT GTACTTGATT TTCACTTTAA AAATTATCAA ATATTGCCAT	2100
	CTCATTTTAA GTATACAAA TGCAAAACAA CCGATTCAAC AGCATATTTT ACACAAGTAA	2160
15	ACCGCTATT TATCAACGTA TATTCGAAGA TGAATTATTT CGATAGTATC TATAGACCAG	2220
	ACGGCATTCT CACTTTTATA GCTATAACTA TACCAGCGTT TTCGTCCTCA AAGGTGCATA	2280
	CTAATAAATC GTAAACATGA CTTTATCAAA TCGTTCCTTC TTGTTAACTA ATTTATCAAA	2340
20	TGTCTCCGGG CCTTTTCTA ACGGTAAAAA ATGAGAAATA ATAGGCTTTA CATTAAATATC	2400
	TTTCGTCTTC ATATAATGTA AGGTTGCCGT CCACTCTTTG CCCGGAAAAT TACTGGACAA	2460
	ACAGTTCCAA GAGCCACATA CTGTCAACTC GTTACGCAGA ATTTTTTCAA AATGAACGCG	2520
25	ATCAATCTCA ATATCATCAT ATGGTATTCC GAGTAATACC ACCTCGCCAC CTTTTTTAGG	2580
	TAGCGTCAAT ATTTGACCAA TCGTAACTTT AGCACCTGAT GATTCTATAG CTAAATCGAT	2640
	TTGATTGGCG TAATGATTTT CGATGAATTT CTCAAGATTT TCTTCTTTTG AATTGATTGT	2700
30	TTGATGTGCG CCCAATGATG TTGCAATATC TAGTTTATGC GCATCTATAT CTATAGCGAT	2760
	GATATGTGCA GCACCAAATA TTCGTGCCCA TTGAATAGCT AACAAACCTA TACTGCCACA	2820
35	CCCCATTACT GCAACAGTCA TACCAGGTTG TATATTCGAT TTATAAAACC CATGCGCAAC	2880
	AACGGCTGAT GGCTCAACCA TTGCTGCTTC AATGTAATCA ACATTGTCTG GAACCTTTAA	2940
	AACATTTTGC GCTGGCAATT TGACATATTC CGCGAACGAT CCAGGTTTAT ATGAGCCAAT	3000
40	GACGAATAAC TTTTCACATC GTGCATATTC ACCTTTTAAA CAATACTCGC ATTGATAACA	3060
	AGGTATTGCT GGGCAACCTG TCACTTTGTC GCCCACATTA ACATGCGTAA CATCACTTCC	3120
	AATGGCATCT ACTACACCTG AAAATTCATG ACCAAATGGC ATACCTTTAA TGTATGGCCC	3180
45	CATTTTTTTG TATCGTGACG TGTCTGAACC ACATATGCCA GTCGCTCGTA CTTTAATAAT	3240
	AACGTCAATC GCACTTTCAA TGAATGGCTT TTCATTATCC TCATACCGTA AATCTTCCAC	3300
	GCCATATAAT TTCAATGCTT TCACTTGTAAT ATCACCTCAA ATTTGATTTA ATTCACAACT	3360
50	TTTTTCTTTT TAAAAATACC TGTGCAAAA TAACCTGCAA TGACAATGGA ATTACTTACG	3420
	AGTAAATGTT CCATATAAAA ATCAGTGATT TGTCTTAATG GCCCAAGCAT AAAAGTTAGC	3480
55		

	TGCTTTAATA CCTTCGCCGG ATTTTAAATG TTGATACGCC TCGTCCCATT TCGAAATATC	3600
	ATATATTTTT GTCACCAAAG CTTCAGCATT TACTAAACCA TCCGCCATAA GTTGCAATGA	3660
5	AGGTTCCCAA TCTGCTGGCT TTTGACTTCT ACTACCAACA ACTGTTATTT CTTTTTGAAT	3720
	CACTTTTTTC ATATCAAATG GAATTTTCAGC ATCCTTAAAA ATACCTATTT GACTGTAGAA	3780
	ACCTTTTTTG CGTAAAATAT CCAAACCTTG TCGTGCTGCT GGAAGTGCAC CTGAACATTC	3840
10	AACAACAACA TCTGCACCGT AACCGTCTGT AATTCCATTG ATATACGTTT TTAAGTCTGT	3900
	TTGTGTGAAA TTGACTACAT AATCCATGTG CAATGCTTCT GCTTTATCTA ATCTGACTTT	3960
15	GTCATTGTCC AATCCAGTTA CCACAACAGT TGCGCCTTTA CTTTTTAACA CTTGTGCTAC	4020
	AAGTAATCCG ATTGGCCCAG GTCCCATTAC AACTGCTACA TCGCCTGAAT TGACTTGAAT	4080
	CTTAGAAACG CCATGATGTG CACATGCTAA TGGTCTGTG ATAGCTGCAG ACTGATACGA	4140
20	TATCGTCTG GAATATGATG CAAACTTTCT TCACGTGCAA TGACATAATT AGTAAATGCG	4200
	CCATCAACTT GTGTTCCAAT ACCTTTTCGA TGGTTGCATA AATTATAGTC TTTTGATTTA	4260
	CAGTATTCAC ACTCATTACA AACATAGAAT GTCGTTTCAG aTGtGACACG GTCACCAACT	4320
25	TTAAAACTT TAACGTCTGC TCCAACCTCA ACGATTCAC CAGAAAATTC ATGACCTAAT	4380
	GTCACTGGAA AATTAACCTT ATAATGACCT TCATAAGTAT GAATATCTGT GCCACAAAT	4440
	CCTGCATAAT GTACTTTAAT CTTTACTTTA TCATCTAGCG GTGTTGCAAC TTCTTTATCA	4500
30	AGAAGTTCTA AGTTGCCATG TCCTTCTCTT GTTTTACTA AAGCTTTCAC CACAAACACC	4560
	TCGATTTTAA ATTGAATAGA CTAAATAGTT TAAAGATAAG ATAGTTAACG ATATTACCAC	4620
	CTTGATCAAT ACTTGAAAT TCAGATGAAC CTTTTGGCAT TTGTACATTC GTACCTTTTCG	4680
35	CCATATCTGT GAAAATGGGT GCTACGTCTG TTGCAATATA TAGTGAAAT GCAATCATAA	4740
	TCGTACCCAC AATGACAGAA TGAATAATGT TTCCTCTTGC TGCACCAACA ATAAACGCGA	4800
40	CAACAAATGG TATCGTTGCT AAGTCACCAA AAGGTAGTAC TTGGTTTCCT GGTAAATAA	4860
	CGGCTAATAA AACAGTGATA GGTACTAAAA TTAATGCTGT CGAAATAACT GCTGGATGAC	4920
	CTAATGCTAC AGCCGCATCC AATCCAATAT AAATTTTCAG TTCGCCAAAA CGTTTATTTA	4980
45	GCCATGTTCT TGCAGACTCT GAACTGGCA TTAAACCTTC CATTAAGATT TTTACCATT	5040
	TAGGCATTAA TACCATTACT GCAGCCATTG ACATTCTTAA ATTAATGATG TCTCCAGGTT	5100
	TGTAACCTGC TAACACACCA ATACCTAAAC CTAAAATTAA GCCGACAAAT ATAGACTCTC	5160
50	CAATGCGCC AAAACGTTTT TGAATTGTTT CAGGATCAGC ATCTAACTTA TTCAGACCGG	5220
	GTACTTTTTG TAACAATTTA ACTAAGTAAA TACCTGGTGC ATAAGAAAT GTACTTCCTG	5280

55



	CTACTTTCAA ACAGATAATT TGGAAAATAA CTGCTGCTAA TAACGCTTGC CAAATACTGC	5400
	CTGATACGGC ATAAACCATT GCTGCTGTAA ACGTATAATG CCAAAAATTC CAAATATCTA	5460
5	CATTCATCGT CTTTGTCACT TTAGTTACTA GCAATACAAC GTTAACTATG ATTCCGAGTG	5520
	GAATAATAAA TGCTGCGACA GATGATGCCC AAGCGATAGA TGATGTTGCT GGCCAACCTA	5580
10	CATCAATCAC ATTCAGACTG ACGCCTAAAT TTTTAACCAT CGCTTGTGCT GCTGGCCCTA	5640
	AATTTTAAAC TAATAAATCG ATGACTAAGA AAATCCCTAC AAAAGCCACA CCTATTGTGA	5700
	AACCAGACCT AAATGCCGCT CCAATTTTCT GCCTAAAGAA TAGGCCAAGC AAGAATATGA	5760
15	CAACCGGTAA AATAACAGTt GCACCTAAAT CTAAAAATCC CTTACAAAA TCAGTGAAGT	5820
	AACTCATATT TAAACCCTCC CTGTTATATA TGCATTGTCA CGATACTTTC CGATTGTGAT	5880
	TACATTTGAC GTTACAGTCA TTTCAACGAC AACCCTTGCT AAATTCGACT GCAGTCCTTT	5940
20	TGAATTACAG tCACTGCGTT TCTATGTCAT CAACAATCAT TTGTCGTGAT AGTCATTTAT	6000
	ATGCAATTTG CATATATTAA TATGTTATCG ACCCACGTTA CATATCAATT CCGTTATTTT	6060
	TGTAACCTCG TTAAGATTTG TTGTTTTGTT TCTTCAATAC CAATACCAGT TAAGAAATTA	6120
25	CGTGCGTTGA TAACTGGGAA TTTATATTCT TTTTTGTCA TTGCAGTTGT AACTAATAAA	6180
	TCTGCAGTGT CTTCATAAGG TCCAACITCT GTAATTTTGA TTTGTTTAAAT ATCTACTTTA	6240
	ATATTGTGTT CCTTTGCCAT TTCTTCAATT GCATTATTTA CTACTGTTGA CGTTGCAATA	6300
30	CCTGCACCAC ACGCTACTAA TACTTGTTC ATTTTCAATT CCTCCAATTA ATTTTTAGTT	6360
	ATATCCAAA TAATCATTGA TTAGTGTGTC TAAAATTGTT TCATCTTTCG TTCGTAGAAT	6420
35	CTGCTCCAAT TTTTCTTAC TTTGAAAAAT TTGCATCAAC TGTGTAACA GCTTAAGTTG	6480
	ATCATCTACT TTATCCATTG CTAACATAAA AACGATTTTC ACTTCTGTCT GTTGATCAAG	6540
	TGTTCCCAT TCAATAAAG GCACTTCTTT TTCTAGAACA GCCACACCTA TCGTTCTATG	6600
40	GTTAATATGT TCGACATCTG TATGCGGTAT AGCGACCGAA CATAGATGCG TTGGTAAACC	6660
	AGTAGCAAAT TCTTTTTCTC TGTCGATGAC TGCATCTTTA AACGTTGACT TCACGAACCC	6720
	ATTTTGAAAT AACACATCTG ACATTTGTGA CAATACGGAT TCTTTATCAG TTGCCGACAA	6780
45	ATTGAGCATT ATATTTTCTT TATGCACTAA TTGCTGTCCC ATCCATTTTC CCTCGCTTCT	6840
	TTATTTGAAT AATTTTTTAA AATCTCATT ACATCAGAAT TTTTGCGACT TTGTATGATG	6900
	CGCTTAATTG CGTCATTGTC TTGCGCCACA TCTCTCAATT GTAGTAACGC TCTTAAGTGT	6960
50	GTCACTTTAT CAACAGCAGC AATAGGTACA ATAATATGGA TTGCTGTGCC ATCTGACATG	7020
	TATATTGGTT CTTGTAATAT CAACATACTC ATCGCTGTTT TATGTACATG CTTTTCAGAG	7080
55		

	TGCATCTCAT GAATATATTT AATATCAATA AAATGATTAG CAACTAACAC ATCACTTGCT	7200
5	TTAGCAATAG CTTCAATCAAT ATTTTCAACA TGATGCATT CTTTCACGTG CCTTGCCGGT	7260
	ATCAAGTCAG CTAAATCTAA TGyCTWATTT tGTGtGACaA TCGATCCATT AATGGTTGAA	7320
	ATTGAATTAT AATTGGCAAT AAAATCTTCT AAACCATCAC GTAGTcTGTA ATGTCATTAA	7380
10	CTGTCGTTGT GCGTTCAATT AATGCCATTA ACTTGTTTAT TTCCTTATCA ATGTCAGCCG	7440
	ATTCCTTATT AATGTACTTC ATCACTTCTT TACGTAACTT TCGTTGCTCA TTTTCAGATA	7500
	AAGCTACTTT TGTGATAAAT AATTTTTTAT GTGTTAGGAC AAACATTGGT GAAAAGACGA	7560
15	TGTCATAATC TAATGTGTAA TTTTCAAATG TTCTAAGTGA AATCGCATCT AAGAAAATAA	7620
	TTTCTGGAAA TAAGTTTCGC AACTCGTATA ACATCATTTG TGATACTGAC GTGCCCTGTG	7680
	TACACACGAT AATAGCTTTT ATCTTGCCAT CGAAGTTTTC ATCTTGACGT CTCAAACCTAC	7740
20	CTCCGAACAA CATGGTTAAA TATGCTATTT CATTATCAGG CAACGATTTT CCGAAATATT	7800
	CAGTTAACGA TTGACATGAT TGTTTCACCA TATGAAATAA GGATTGATAA TTTCTTGTA	7860
	AAGGATTTAT TAAITCATCA CGATCCGTTA AGTTATATTT AATCCTATAA AAAGCAGGCG	7920
25	TTAAATGTAA CAAGAGTTGC TGTGATAATT TCTCCTTATC TTCAATGTTA ATAAAAGTGA	7980
	TTTGTTCAAA ATGGTGAATC ATTTGAGCGA TGGCCATCGT TAAATTCGAT ATGCTATCTG	8040
	ATTCTTGCAA ATCAGTCCAT TGCACACTTG TTGAAAGTAA GTGTAATGTC AAATATAACT	8100
30	TTTCCGCTTC TGGCAAATCC GGCTCATGTT GCGTCATAAT CTCCGTTGCT TGATATTCTT	8160
	TCGTATCCCT CAAATACTGA TAATTAATAT TTAATGGATT CATCACATGA CCACTTTGAA	8220
35	TTCGTCTACG AATCACACAA AGGACATAAG GCAATGAACT AAGTGATTTG TCTATAAAGC	8280
	GACTCTTCAA AAATGTTTCT ACCTGTTTGA TCTTGCTTTT TTGATATGCG ATATCTTCGA	8340
	ATGTTAAGTT GAGCGCCTTT AAAACTTCAC TTTTAGTAAT ATCATGATTC AACCTTTGAT	8400
40	CAATCAACTT AATGAAGAAA CGGCGAACTT CAAATTCATC ACCAACAATT TCATAACCAT	8460
	GTTTTGAGA ATACTTAAGT GACAAACCAT GATTTTCCAA TTGCTCTTTC ACATGATTTA	8520
	TATCGTGAAT GACAGTATTT TTAGTGACTT GTAAATCAAT TGAAAAATGG TTTAGAGACA	8580
45	TTGCGTTTTT CTTACTAAAA AGCATGAGCA TTAAATAATA ACGACGTGTT TCTATGCTAA	8640
	AAATGACATT GTTGCCGTTT AACATTTGCT GCTCCGATAC ATCTCGCTTG AATAACGTCA	8700
	TGATTTGAGA ACTTACAATA AAATTTCTT GGCTTGTTCT TTCAAGTTTT GGATAACCCT	8760
50	CTTGTTCAAG CCACAAATTG ATTTTTTGAA TGCGATATCC TAGTTGTCTA CGAGACAAAC	8820
	CAAATATCGA TTCAAGTTCT TTACCATGAA TAGTAGGATT CAATACAATT TCTCTGAGTA	8880

55

TCAATCGTCA CACCGATGTA CACACTTTGA ACACATATTT TCAAAATGAG CATGTACATC 9000  
 ATTGTGATGT TTTAACAACA TTTCAATTAT ATCTATATTT TTTGTGATTT TAATCTTTTA 9060  
 5 AAATAAAGCA ATTGAAATTT TTGCATATAT TTTTGTGTTT TGTGTTTTTT TGAAGCATTT 9120  
 TTAACATACA TATCTCAATC ATTATCAAAT TGTATGACC ATTGTAACCC AATACAAAAA 9180  
 10 CCCTAAGGAC GCTTATATCA GCGCCTTAG GGTAACTGT ATCTATTTAA TTAAGTATTA 9240  
 TTATTCGTAT GTACGTAAC TATGGTCTAT CAAGTTCCAC ACTTCTTCAA CATCAACTGC 9300  
 TGTAGCAAAA TAAGCATTGG CAGGCTTACC TGTAACATGA TTTAAATCGA CAGCCATAGT 9360  
 15 GCCATAAGTT AGTGGACTTT GATGTTCAAT GTCGATATTA ACGGGTACCA TTGTAAACAA 9420  
 TTCTGGTTGT AACAAATACA AAATTGTACA AGCATCATGT ATTGGACCAC CATCCATATT 9480  
 AAAGTGAGTC TTGTATGTCT TCTTAAAGAA TTGCAATAAT TCTACGACGA ACTGTGCAAC 9540  
 20 AGGATTATTG ATACTTTCAA AGCGTTCAAT CACGTGATCG TCGGCTAAAA CTTGATGTGT 9600  
 TACATCTAAA CCAAACACAT TTATAGTAAT CCCACTTTCA AAAACACGCT TCGCTGCTTC 9660  
 AGCATCTACC CAAATATTGA ATTCTGCTGT AGGCGTCCAA TTTCCAAATG TACCACCACC 9720  
 25 CATCAAAGTA ATAGATTCAA TATGCTCAGC GATTCTTGGC TCACGAATCA ATGCCGTTGC 9780  
 TACATTCGTA AGAGGACCTG TCGCTACAAT TGTTACAGGT GTATCACTCG TCATCACTTT 9840  
 GTTTATAATC ACATCTGATG CTGGCATTGC AACTGCTTGA CGTGATGGTG TCGACGGTAG 9900  
 30 TTTCCGACCA TCTAATCCAG ATTCCCATG TATTTTCAGAA GCAAAGGCAG CTGGTTTAAT 9960  
 TAACGGCCTA TCCGCACCTT TCGCTACTGC TATATCTTGG CGTCCCATAA TATCCAATAC 10020  
 GTTCAAGGCG TTTGTCGTAT TCTTGTCAAC TGATTGATTA CCTGCGACTG TTGTTACAGC 10080  
 35 TAATATCTCT AGTGGACTGT CAATTGCCCC CGCTAAAATT AATGCTATTG CATCATCGTG 10140  
 TCCTGGATCA CAATCCATAA TAATCTTTCT TTTCAATTAT ATATCCACCT TTCTTAAGTT 10200  
 GTTATCGATA GCTTATGTAT ATTTATTTAT GTGGTGAATC ATGTTTATTT TGAAAAATAG 10260  
 40 TTTTAACTTT CTCATATTTT TGGATACAAA CACTATTTAT CTATTTTATG GCTTATAAAT 10320  
 TTATCCGATA TGCCTTATCA ACCTACCTCG CTAAAAATAG GATGTCTACA TATCTATACC 10380  
 GACTTTTGTC AACTCATTTT CACAACAATA TAAACAGCAA TTTATATGAT TGTACATGA 10440  
 TTCAAACAAT TTTTATGAAA AATATTTTCA TACACAGAAT ATATATTGAT ATTAAATTTT 10500  
 TCAAAAGCTA TATTGAGAAT AATTAGGAGG GATGTTGATG AAATCTTTAT TTGAAAAGC 10560  
 50 ACAGCAGTTC GGCAAGTCCT TTATGTTACC TATCGCAATC TTACCAGCTG CAGGTCTATT 10620  
 GTTGGGTATC GGTGGTGCAT TAAGTAATCC AAACACCGTT AAAGCATACC CTATTTTAGA 10680

55

## EP 0 786 519 A2

AAATTTACCG GTCATCTTTG CAATTGGTGT CGCAATCGGA TTATCTAGAA GCGATAAAGG 10800  
 TACTGCAGGT tTAGctGCGC TGCTCGGTTT CTTAATTATG AACGCAACTA TGAATGGCTT 10860  
 5 ATTAACATATC ACGGGCACAT TGGCAAAAGA TCAGCTTGCA CAAAATGGAC AAGGCATGGT 10920  
 GCTCGGTATA CAAACGGTTG AAACCGGTGT TTTTGGCGGG ATTATCACAG GTATTATGAC 10980  
 CGCAATACTT CACAACAAAT ATCACAAAGT GGTATTACCA CCGTATTTAG GTTCTTTGG 11040  
 10 TGGCTCTAGA TTTGTCCCTA TTGTCACAGC ATTTGCCGCA ATCTTTTtag GTGTATTGAT 11100  
 GTTTTTTCATT TGGCCAAGCA TACAAGCCGG CATTTATCAT GTTGGTGGAT TTGTAACGAA 11160  
 AACAGGTGCC ATCGGTACTT TTGTTTATGG CTTTATCTTA AGATTGTTAG GTCCACTCGG 11220  
 15 TTTACACCAT ATTTTTTACT TACCGTTTTG GCAGACGGCA CTTGGTGGTA CTTTAGAAGT 11280  
 CAAAGGGCAC TTAGTTCAAG GTACGCAGAA CATCTTCTTT GCTCAACTTG GTGATCCAGA 11340  
 TGTGACGAAG TATTATTCAG GTGTGTCACG CTTTATGTCA GGCCGTTTTA TTACGATGAT 11400  
 20 GTTCGGCTTA TGTGGTGCCG CACTTGCAAT TTATCACACA GCTAAACCTG AACATAAAAA 11460  
 AGTTGTCGGC GGTTTAATGT TATCCGCTGC ACTCACTTCA TTTTAAACAG GTATTACCGA 11520  
 25 ACCTTTAGAG TTTAGTTTCT TGTGTGTCG ACCTATCTT TATGTAATCC ATGCCTTCTT 11580  
 TGATGGATTA GCATTTATGA TGGCAGACAT TTTCAACATT ACAATTGGTC AAACCTTCAG 11640  
 TGGAGGCTTT ATCGATTTCT TACTCTTTGG TGTGCTACAA GGTAATAGTA AAACAAACTA 11700  
 30 CCTATACGTC ATACCTATTG GAATTGTGTG GTTCTGTTT TATTACATCG TTTTCAGATT 11760  
 CTTAATTACG AAATTTAATT TCAAAACACC TGGTCGAGAA GATAAAGCTG CAGCACAACA 11820  
 AGTTGAGGCT ACTGAAAGAG CACAACTAT TGTGCTGGT TTGGGAGGCA AAGATAACAT 11880  
 35 TGAAATCGTT GACTGTTGTG CAACGAGACT ACGCGTCACA CTTTATCAA ATGACAAAGT 11940  
 CGATAAAGTA TTAATCGAAA GTACTGGTGC CAAAGGTGTA ATCCAGCAAG GCACTGGTGT 12000  
 GCAAGTAATT TATGGGCCTC ACGTTACAGT TATCAAAAAT GAAATTGAAG AATTGCTCGG 12060  
 40 GGATTAAGAC TAACCGAAAT ATCAACAGAA CTAATGGCAA CGATGTACGA AGTAAGAAGT 12120  
 GACATCGTTG CTTTATTTT TAATGTTACA TTTGAAGCAT TAAGTTCATC ATGCACTGTA 12180  
 GTGAGCCCGC AAATCGCCTC TGCTAGACAA TCATCTTAAT GCTATGATTA AAGCTTAAGT 12240  
 45 GCCAGATTG AATTTAATTT CAACAACGAC TTTCACTACA TTAAAAATAG GGCCACTCGA 12300  
 CACATATAGT TGTATCAAAT AGCCCTTTAT ACAATTTTTT GGGTAAGGTT TTACAATTTT 12360  
 50 TGGGATGGTA TAGATTTTAT AAAAAGTTAT TTAAGTCTT CTGCTTCAGC CATAATATCT 12420  
 TTTAATGTTT TAGCTGAATG TGCGAACTTG CTTTGTCTT CGTCGTTTAA TGGGATTTCT 12480

55

	TCCTCATATT	CGCCTTCTAA	TAATGCTGAT	ACAGTCAATA	CGGCATCTTC	ATTTCTGAAA	12600
	ATCGCTTCAG	TAATTCTAGC	TAATCCCAT	GCAACACCAT	AATAAGTGGC	ACCTTTAGCT	12660
5	TGAATAATGT	CATATGCTGC	ATCACGTGTT	TGAACAAAAA	TTTGTTCAAT	TTGCGCTTTG	12720
	CCCTCAGGAC	GTTGTTCAAG	TAATGTCTTC	AAAGGTTGAC	CCGCAATATT	AGCGTGTGAC	12780
	CATACTGGTA	ATTAGTGTG	ACCATGTTCA	CCAATAATTT	GAGCATCGAC	GCTACGTGGC	12840
10	GCAACATCGn	AcgyTcGCTT	AACAATAATC	TAAAGCGTGC	AGAGTCTAAA	ATTGTACCAG	12900
	AACCTATAAC	ACGTTCTTTA	GGTAAACCAG	AGAATTTCCA	TGTGCATAC	GCTAAAATAT	12960
15	CAACAGGATT	TGTAGCTACC	AAGAAAATAC	CATCAAATTT	TGATGCCATT	ACTTCACCAA	13020
	CAATTGATTT	GAATATTTTC	AAGTTTTTAG	ATACTAAATC	TAAACGTGTT	TCTCCAGGTT	13080
	TTTGTGCAGC	ACCAGCACAG	ATGACAATA	GATCCGCATC	ATGACAATCA	CTGTATTGCG	13140
20	CAGCTTTCAC	ACGAACTGTT	GTTGGAGAAT	ATGGTGTGGC	ATGTTTTAAA	TCCATAACAT	13200
	CTCCTCGAAC	TTTTTCAGTG	TCTAAATCAA	TGATGACTAA	TTCATCAACA	ATGCTTTGGT	13260
	TCACTAATGA	AAATGCGTAG	CTTGAACCTA	CTGCACCATT	ACCTATTAAT	ACAACTTTGT	13320
25	TCCCTTTAAA	TTTGTTCAAT	ACAAAACTC	CCTTATGATT	AATTCACTAA	CATACATGTA	13380
	GCTTCAAATA	TGTTAGTTTA	ATGCTGCTTA	TTGACGATAC	AAAAGCAAAT	AAACATCTCT	13440
	TTTATTTTCA	ACGCATAACT	TAAAAGGTCA	TGTGTCATCC	GCTTTTAAGT	TTGTGATTTA	13500
30	TTTCACATAT	AAAATGTAAC	ATGCATTAAG	TACTGGGTCA	ATATTAAATT	GTGATTTATT	13560
	TCACATTTTA	TTTAATTTT	TACACCTTTT	TAATTGTAT	mCGATTACAT	CTTAGATGTC	13620
	TTTAGTCTTC	GTAATTCGCC	AGTGATTATT	TACACTTTCA	CATTTTATT	ATCATGTTTA	13680
35	CTTTTTTCTA	GGAAAACAAC	AATGTTTTTT	GAATTAGTCA	AATAAATGCG	CTCAATCGTC	13740
	GGTGTCGAAA	CAGACAATTG	TACACAATGC	TTATTGATAA	GTATTTAAAA	AATTAAAAAT	13800
40	GTCATACAAT	TATCAAATTT	GCCATTTTAT	TTATATTTTC	TCAAACCAAT	TAATTGAATA	13860
	TCGAAATTTT	TAGTAGAATA	ATCAAATAT	ACAGATTAAA	GGAGGAGTAT	CATGCTTACA	13920
	GAACAAGAGA	AAGACATTAT	CAACAAACG	GTGCCTTTAC	TTAAAGAGAA	AGGGACAGAA	13980
45	ATTACGTCAA	TCTTTTATCC	AAAAATGTTT	AAAGCGCATC	CTGAACTTTT	AAACATGTTT	14040
	AATCAAACGA	ACCAAAAACG	AGGCATGCAA	TCTTCAGCAT	TAGCACAAGC	TGTAATGGCC	14100
	GCAGCGGTTA	ATATCGATAA	CTTAAGTGTT	ATTAAACCAG	TCATTATGCC	AGTCGCATAT	14160
50	AAACACTGCG	CACTACAAGT	TTATGCTGAA	CATTATCCAA	TTGTGGGGAA	AAATTTATTA	14220
	AAAGCCATTTC	AAGACGTGAC	AGGATTAGAA	GAAAAATGACC	CTGTCATTCA	AGCTTGGGCA	14280

55

## (2) INFORMATION FOR SEQ ID NO: 58:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8779 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

GGTATTTTnG GAnGGGTACC TAAAGCAATT CCGGCAAAGG GTnAATCCAG GTACCGAAAT 60  
 GGACTTCCCG TTATCGATAA TACCGACATA TATTGTGACA AGTAGATTTT ATGGACATTT 120  
 AGGCTTACTT TTACTTGTGA TAATTGCATG TATGTTTACT GGTATTTAtC CaTCaATACA 180  
 TATCATTCAA TTATTGATAT ATGTACCGTT TTGTTTTTTC TTAAGTGCCT CCGTGACGTT 240  
 ATTAACATCA AACTCGGTG TGTTAGTTAG AGATACACAA ATGTTAATGC AAGCAATATT 300  
 AAGAATATTA TTTTACTTTT CACCAATTTT GTGGCTACCA AAGAACCATG GTATCAGTGG 360  
 TTTAATTCAT GAAATGATGA AATATAATCC AGTTTACTTT ATTGCTGAAT CATACCGTGC 420  
 AGCAATTTTA TATCACGAAT GGTATTTTCA GGATCATTGG AAATTAATGT TATACAATTT 480  
 CGGTATTGTT GCCATTTTCT TTGCAATTGG TGCGTACTTA CACATGAAAT ATAGAGATCA 540  
 ATTTGCAGAC TTCTTGTAAT ATATTTATAT GACGAAACCC CGCTAACCAT TAATAAATGG 600  
 AAGTGGGGTT CATTTTGTGTT TATAATTTAA GTAAATAACA TATTAAGTTG GTGTATTATG 660  
 AACGTTTTTAA TAAAGAAATT TTATCATTTG GTAGTTTCGA TACTTTCTAA AATGATTACG 720  
 CCTCAAGTGA TTGATAAACC GCATATCGTA TTTATGATGA CTTTTCAGAG AGATATTAAG 780  
 CCTATCATCA AAGCATTAAA TAATTCGTCG TATCAGAAAA CTGTTTTAAC AACACCAAAA 840  
 CAAGCGCCTT ATTTATCTGA ACTTAGCGAC GATGTTGATG TGATAGAAAT GACTAATCGA 900  
 ACATTGGTAA AACAAATTAA GGCTTTGAAA AGCGCGCAGA TGATTATTAT CGATAATTAT 960  
 TACCTATTGC TAGGTGGATA TAATAAGACT TCTAATCAAC ACATTGTTCA AACGTGGCAT 1020  
 GCAAGTGGTG CATTAAAAA CTTTGGCTTA ACAGATCATC AAGTCGATGT GTCTGACAAG 1080  
 GCAATGGTTC AGCAGTACCG TAAAGTTTAT CAAGCGACGG ATTTTACTT AGTGGGTTGT 1140  
 GAACAAATGT CACAATGTTT TAAACAGTCT TTAGGTGCAA CAGAAGAGCA AATGCTGTAT 1200  
 TTTGGGCTTC CGAGAATTAA TAAATATTAC ACAGCTGATA GAGCAACGGT TAAGGCAGAG 1260  
 TTAAAGGATA AATATGGAAT TACAAATAAG TTGGTATTAT ATGTACCAAC ATATAGAGAA 1320  
 GATAAAGCAG ATAATAGGGC TATTGATAAA GCTTATTTTG AAAAATGTTT ACCAGGATAT 1380

	ATCGACACGT CTACATTAAT GCTAATGTCA GATATAATTA TTAGCGACTA TAGTTCGCTG	1500
	CCAATAGAAG CTAGCTTGTT AGATATTCCA ACTATATTTT ATGTGTATGA TGAAGGAACA	1560
5	TATGATCAGG TGAGAGGCCT GAATCAATTT TACAAAGCAA TACCGGATAG CTACAAAGTG	1620
	TATACTGAAG AAGATTTAAT AATGACGATA CAAGAAAAAG AACATCTATT AAGTCCGTTA	1680
	TTTAAAGATT GGCATAAGTA TAATACTGAT AAAAGTTTAC ATCAGCTCAC AGAATATATA	1740
10	GATAAGATGG TGACAAAATG AGGTTTACGA TAATCATACC TACATGTAAT AATGAGGCAA	1800
	CAATTCGACA ATTGTTAATA TCTATTGAGA GTAAAGAACA CTATAGAATC CTTTGTATTG	1860
	ATGGTGGTTC TACTGATCAA ACAATTCCTA TGATTGAACG GTTACAAAGA GAACTCAAGC	1920
15	ATATTTTCATT AATACAATTA CAAAATGCTT CGATAGCTAC GTGTATTAAAT AAAGGTTTGA	1980
	TGGATATCAA AATGACAGAT CCACATGATA GTGACGCATT TATGGTCATA AAACCAACAT	2040
20	CAATCGTATT GCCAGGTAAA TTAGATAGGT TAACTGCTGC TTTCAAAAAT AATGATAATA	2100
	TTGATATGGT AATAGGGCAG CGAGCTTACA ATTACCATGG TGAATGGAAA TTGAAAAGTG	2160
	CTGATGAGTT TATTAAAGAC AATCGAATCG TTACATTAAC GGAACAACCA GATTTGTTAT	2220
25	CAATGATGTC TTTTGACGGA AAGTTAATCA GTGCTAAATT TGCTGAATTA CAGTGTGaCG	2280
	AAACTTTAGC TAACaCATAC AATCACGCAA TACTTGTCaA GGCGATGCAA AAAGCTACGG	2340
	ATATACATTT AGTTTCACAG ATGATTGTGCG GAGATAACGA TATAGATACA CATGCTACAA	2400
30	GTAACGATGA AGATTTTAAT AGATATATCA CAGAAATTAT GAAAATAAGA CAACGAGTCA	2460
	TGGAAATGTT ACTATTACCT GAACAAAGGC TATTATATAG TGATATGGTT GATCGTATTT	2520
	TATTCAATAA TTCATTAAAA TATTATATGA ACGAACACCC AGCAGTAACG CACACGACAA	2580
35	TTCAACTCGT AAAAGACTAT ATTATGTCTA TGCAGCATTG TGATTATGTA TCGCAAAACA	2640
	TGTTTGACAT TATAAATACA GTTGAATTTA TTGGTGAGAA TTGGGATAGA GAAATATACG	2700
40	AATTGTGGCG ACAAACATTA ATTCAAGTGG GCATTAATAG GCCGACTTAT AAAAAATTCT	2760
	TGATACAACCT TAAAGGGAGA AAGTTTGCAC ATCGAACAAA ATCAATGTTA AAACGATAAC	2820
	GTGTACATTG ATGACCATAA ACTGCAATCC TATGATGTGA CAATATGAGG AGGATAAATT	2880
45	AATGAAACGT GTAATAACAT ATGGCACATA TGACTTACTT CACTATGGTC ATATCGAATT	2940
	GCTTCGTCGT GCAAGAGAGA TGGGCGATTA TTTAATAGTA GCATTATCAA CAGATGAATT	3000
	TAATCAAATT AAACATAAAA AATCTTATTA TGATTATGAA CAACGAAAAA TGATGCTTGa	3060
50	ATCAATACGC TATGTCGATT TAGTCATTCC AGAAAAGGGC TGGGGACAAA AAGAAGACGA	3120
	TGTCGAAAAA TTTGATGTAG ATGTTTTTGT TATGGGACAT GACTGGGAAG GTGAATTCGA	3180

55

	TAAATCAAA CAAGAATTAT ATGGTAAAGA TGCTAAATAA ATTATATAGA ACTATCGATA	3300
	CTAAACGATA AATTAACITA GGTATTATA AAATAAATAT AAAACGGACA AGTTTCGCAG	3360
5	CTTTATAATG TGCAACTTGT CCGTTTTTAG TATGTTTTAT TTTCTTTTC TAAATAAACG	3420
	ATTGATTATC ATATGAACAA TAAGTGCTAA TCCAGCGACA AGGCATGTAC CACCAATGAT	3480
	AGTGAATAAT GGATGTTCTT CCCACATACT TTTAGCAACA GTATTTGCCT TTTGAATAAT	3540
10	TGGCTGATGA ACTTCTACAG TTGGAGGTCC ATAATCTTTA TTAATAAATT CTCTTGATA	3600
	GTCCGCGTGT ACTTTACCAT CTTGACTAC AAGTTTATAA TCTTTTTTAC TAAATCACT	3660
	TGGTAAACA TCGTAAAGAT CATTTTCAAC ATAATATTTT TTACCATTTA TCCTTGCTC	3720
15	ACCTTTAGAC AATATTTTTA CATATTTATA CTGATCAAAAT GAGCGTTCCA TTAATGCATT	3780
	CCCCATCATA TTACGTTGCT TCTCGCCACC AAGGTTTTTA TAGTCTCCTG CACCCATGAT	3840
	AACCTGATTA ATTCTAAAT TACCTCGTTT GGTAAGTAAT GTATGGTTGT AATTGCTGT	3900
20	ATCACTTGAT CCAGTTTTTA AACCATCTGT ACCCGGCAAA CTCATTTTGT CACCTTCCAA	3960
	TGAAAAGTTG AATGTGTAAT ACGTAACTGC ATGCGTTGTT GGTGCTAACT GCTTTGTAAA	4020
25	GTCTAATATT TTAGGTGTCT CTTTAATCAC GTGTAAATCT AAAATGGCAT AGTCTCTAGC	4080
	AGTCGTTACA GTACGTTCTT GGTCTTTATA CTTTGTTGGT GCAAATGTAC GTAATCTTGA	4140
	ATTTTCAGCA CCCGTTGGAT TGACGAAATG TGTATTTTTT ATTCCGATAG CTTTAGCTTT	4200
30	GTTATTCATT AAATCAACGA AATCGCTGGT GTTTTTTGAA ACCTTCTTAG CTAAATTA	4260
	TGCCGCGGCA TTAGTAGAAT TAGATACTGT AATTGTAAAT AGGTCTGCGA TTGTCCATAC	4320
	TTGTCCAGGA TATAGTTTCG TATTACTCAA CTCAGGTAGT GTAGACATAA TATATTCTTT	4380
35	GTTTCGTCATT GTGACTGTGT CATCAAGTGA AAGCTGCCCC TTATTTACAG CTTCCAATGT	4440
	TAAGTACATT GTCATTAATT TAGTCATAGA CGCTGGATTC CACTTAGTAT CGATATTGTA	4500
	TTGATACAGT AATTGTCCAG TTTGACTTAC ATTAACAGCA CTCGTCGGTT CGTATGCAGC	4560
40	CGACAAACCT GCATAACCAT ATTGATTGTC TGCTTGATCA GGGGTTACGT CACTGTTAGT	4620
	AGCTTGTCAT TATGGTGTCA TAATACTTAA TGTAAACAT AAAATGATGA TAATAGATAT	4680
45	TAAATTTTTT ATAAAGCGTT AATCTTCCCT TTTCCAATTC TTAATATTC CCTAAAAGCA	4740
	ATGGTTATTC CTAATTACGG AAATCATGTC TAATTCACCT CACCTTAATT AAATTGTTGA	4800
	AAATAAGTT TTCTGCAGTT AATTTGAAAA ATAATGCAAA TATATTACGT GTGTAGCTAA	4860
50	AGGTGTTATA ATGTTTGTAC GAAGAGCAAA CTTACTCAAA AGCGATTAAT TTTCATGTTT	4920
	TAATATAAAG ACTTTGAGAA GTTATTACAA AAAATGCAAT AGAAATATTC TATCATATAA	4980

55



AAGTATATGA TAGAAATGCA TGTATCTATC TAAATGAATT AACTATAAAT TTCAAACAGA 5100  
 AGAGGTAAAA CTATGAAACG AGAAAATCCA TTGTTTTTCT TATTTAAAAA ACTATCATGG 5160  
 5 CCAGTGGGTC TTATCGTTGC AGCTATCACT ATTTCATCAC TAGGGAGCTT AAGTGGACTA 5220  
 TTAGTGCCAC TGTTTACTGG ACGAATTGTA GATAAATTTT CCgTGAGCCA TATCAATTGG 5280  
 AATCtAATCG CATTATTTGG TGGTATCTTT GTCATCAATG CTTTATTAAAG CGGATTAGGT 5340  
 10 TTATATTTAT TAAGTAAAAT TGGTGAAAAG ATTATTTATG CGATACGCTC AGTTTTATGG 5400  
 GAGCATATCA TACAATTAAA AATGCCATTG TTTGACAAAA ATGAAAGTGG TCAATTAATG 5460  
 15 AGTCGATTAA CTGACGATAC GAAAGTGATA AATGAATTTA TTTACAAAA GCTACCTmAC 5520  
 TTATTACCAT CAATCGTTAC ATTAGTTGGG TCACTAATCA TGTTATTTAT TTTAGATTGG 5580  
 AAAATGACAT TATTAACATT TATAACGATA CCGATATTGCG TTTTaATTAT GATTCCTCTA 5640  
 20 GGTCGTATTA TGCAAAAGAT ATCGACAAGT ACACAATCTG AAATTGCAAA CTTCACTGGT 5700  
 TTGTTAGGGC GTGTCCTAAC TGAAATGCGT CTTGTTAAAA TATCAAATAC AGAGCGTCTT 5760  
 GAATTAGATA ATGCACATAA AAATTTGAAT GAAATATATA AATTAGGTTT AAAACAGGCT 5820  
 25 AAAATTGCGG CAGTTGTACA ACCAATTCA GGTATAGTTA TGTTGCTAAC AATTGCAATT 5880  
 ATTTTAGGTT TTGGTGCAAT AGAAATTGCG ACTGGTGCAA TCACTGCAGG TACATTAATT 5940  
 GCAATGATAT TTTATGTTAT TCAGTTATCT ATGCCTTTAA TCAATCTTTC CACGTTAGTT 6000  
 30 ACAGATTATA AAAAGGCAGT CGGTGCAAGT AGTAGAATAT ACGAAATCAT GCAAGAACCT 6060  
 ATTGAACCGA CAGAAGCTCT TGAAGATTCT GAAAATGTAT TAATTGATGA CGGTGTATTG 6120  
 TCATTTGAAC ATGTAGACTT TAAATATGAT GTGAAGAAAA TATTAGATGA TGTGTCGTTC 6180  
 35 CAAATCCAC AAGGTCAAGT GAGTGCTTTT GTAGGCCCTT CTGGGTCTGG TAAAAGTACG 6240  
 ATATTTAATC TGATAGAACG TATGTATGAA ATTGAGTCAG GTGATATTAA ATATGGCCTT 6300  
 40 GAAAGTGTCT ATGATATCCC GTTATCTAAG TGGCGACGCA AAATTGGATA TGTTATGCAA 6360  
 TCAAATTCGA TGATGAGTGG TACAATTAGA GACAATATTT TATACGGAAT TAATCGTCAT 6420  
 GTTTCAGATG AAGAACTTAT TAATTATGCT AAATTAGCGA ACTGTCATGA TTTTATCATG 6480  
 45 CAATTTGATG AAGGATATGA CACGCTTGTA GGTGAACGAG GATTGAAACT GTCTGGCGGA 6540  
 CAACGTCAAC GTATTGATAT TGCTAGAAGT TTTGTTAAAA ATCCTGATAT TTTGTTACTT 6600  
 GATGAAGCAA CAGCTAATCT CGATAGTGAA AGTGAATTGA AAATTCAAGA AGCTTTAGAA 6660  
 50 ACATTGATGG AAGGTAGAAC AACGATTGTC ATTGCGCATC GTTTGTCTAC AATTAAAAAA 6720  
 GCCGGTCAAA TTATATTCTT AGACAAAGGA CAGGTAACAG GTAAAGGTAC GCATTAGAA 6780

55

	TTTTATATAT ATAAGTAAGC TTGGAGCAAA TACACATATA CCATCGAGGA AATTAAAGTG	6900
	TGGCACATTG ATGGATATAG ATGTTAATAA ATTGCTTCAA GCTTTTGTCT ATTTTAAATC	6960
5	ATTTGAGAAG TTACGACATA ATAATTCCTTA AATTAATGAA ATCGATATTT TAAGAAAAAA	7020
	ATGCTCATGG TATAATACAA GTTATAAGCA AACATACATA TATTAAATAC TGTAGCCACG	7080
	AGTCATAATT CTTCATATTT TACATAGCAA TTAACTGAT TTTAGAGTCC ACGGTACAGA	7140
10	AGTTTGATAT TTCAATGTTT CTAAATTTTT AAAAATTAA ATCATAGGTG GGTGCCAAAT	7200
	GTTTTTATTA ATCAACATTA TTGGTCTAAT TGTATTTCTT GGTATTGCGG TATTATTTTC	7260
	AAGAGATCGC AAAAATATCC AATGGCAATC AATTGGGATC TTAGTTGTTT TAAACCTGTT	7320
15	TTTAGCATGG TTCTTTATTT ATTTTGATTG GGGTCAAAA GCAGTAAGAG GAGCAGCCAA	7380
	TGGTATCGCT TGGGTAGTTC AGTCAGCGCA TGCTGGTACA GGTTTTGCAT TTGCAAGTTT	7440
20	GACAAATGTT AAAATGATGG ATATGGCTGT TGCAGCCTTA TTCCCAATAT TATTAATAGT	7500
	GCCATTATTT GATATCTTAA TGTACTTTAA TATTTTACCG AAAATTATTG GAGGTATTGG	7560
	TTGGTTACTA GCTAAAGTAA CAAGACAACC TAAATTCGAG TCATTCTTTG GGATAGAAAT	7620
25	GATGTTCTTA GGAAATACTG AAGCATTAGC CGTATCAAGT GAGCAACTAA AACGTATGAA	7680
	TGAAATGCGT GTATTAACAA TCGCAATGAT GTCAATGAGC TCTGTATCGG GAGCTATTGT	7740
	AGGTGCGTAT GTACAAATGG TACCAGGAGA ACTGGTACTA ACGGCAATTC CACTAAATAT	7800
30	CGTTAACGCG ATTATTGTGT CATGCTTGTT GAATCCAGTA AGTGTGAAG AGAAAGAAGA	7860
	TATTATTTAC AGTCTTAAAA ACAATGAAGT TGAACGTCAA CCATTCTTCT CATTCTTTGG	7920
	AGATTCTGTA TTAGCAGCAG GTAAATTAGT ATTAATCATC ATCGCATTG TTATTAGTTT	7980
35	TGTAGCGTTA GCTGATCTAT TTGATCGTTT TATCAATTTG ATTACAGGAT TGATAGCAGG	8040
	ATGGATAGGC ATAAAAGGTA GTTTCGGTTT AAACCAAATT TTAGGTGTGT TTATGTATCC	8100
	ATTTCGCTA TTA CTGGTT TACCTTATGA TGAAGCGTGG TTGGTAGCAC AACAAATGGC	8160
40	TAAGAAAATT GTTACAAATG AATTTGTTGT TATGGGTGAA ATTTCTAAAG ATATTGCATC	8220
	TTATACACCA CACCATCGTG CGGTTATTAC AACATCTTA ATTTCAATTG CAAACTTCTC	8280
45	AACGATTGGT ATGATTATCG GTACATTGAA AGGCATTGTT GATAAAAAGA CATCAGACTT	8340
	TGTATCTAAA TATGTACCTA TGATGCTATT ATCAGGTATC CTAGTTTCAT TATTAACAGC	8400
	AGCTTTGCTT GGTTTATTTG CATGGTAATA TGTGGAAGAG TGA CTATGAT AATACATTTT	8460
50	AACTAATAAA TATGTCCAGG CATGTCGTCT ATTGATATAG GTGAGATGCT TGGACTTTTT	8520
	TATTATTGAT ATAAAGGTAT nTAAATATTT TTAAAGTTAC CGAAATTGAA GCATTATAAA	8580

55

GACAGTAAGG ACTAGGTACA GTCATAGTAC TTCGAGCAAA ATTTGTTTTG TTATTATAAA 8700  
 CAACACAAAG GAGATAACTT CTCTAaTGAA GAAGTTAAAA ACATTATAGC AGACAATGAA 8760  
 5 ATGAAAGTAA ATTAAAAAT 8779

(2) INFORMATION FOR SEQ ID NO: 59:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 31096 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

GTTGCAGTAG TCAAAGAATT AAACAAGGTG AAGGcGTGTA GCTTGcACAC CCGAAAATGT 60  
 20 GCGTAAGTTA aCGGATGCAG GACATAAAGT AATTGTTGAA AAAAATGCTG GCATTGGTTC 120  
 AGGATTTTTCT AACGATATGT ATGAAAAAGA AGGCGCTAAG ATCGTAACTC ACGAACAAGC 180  
 ATGGGAAGCT GATCTTGTTA TCAAAGTAAA AGAACCTCAT GAAAGCGAAT ATCAATATTT 240  
 25 CAAAAAGAAT CAAATTATCT GGGGATTTTT ACATCTAGCA TCTTCAAAAG AAATAGTAGA 300  
 AAAAATGCAA GAAGTTGGTG TAACTGCGAT TAGTGGTGAA ACCATTATAA AAAATGAAAA 360  
 AGCAGAATTA TTAGCGCCAA TGAGTGCTAT AGCAGGTCAA CGCTCAGCAA TTATGGGAGC 420  
 30 TTACTACTCT GAAGCACAAC ATGGTGGTCA AGGTACTTTA GTGACTGGTG TACATGAAAA 480  
 TGTGGATATA CCTGGTAGTA CATATGTGAT TTTCGGTGGT GGAGTAGCAG CAACAAATGC 540  
 AGCAAATGTT GCCTTGGGAC TAAATGCTAA AGTAATCATT ATCGAGTTAA ACGATGACCG 600  
 35 CATTAAATAT CTTGAAGATA TGTATGCAGA AAAAGATGTC ACAGTAGTCA AATCAACACC 660  
 AGAAaATTTA GCAGAACAAA TTAAGAAAGC AGATGTATTT ATTTCTACAA TTTTAATTTc 720  
 AGGTGCGAAA CCGCCAAAAT TGGTTACTCG TGAGATGGTT AAATCAATGA AAAAAGGTTC 780  
 40 AGTATTAATC GATATAGCTA TTGACCAAGG TGGAACTATT GAAACAATTA GACCAACTAC 840  
 AATTTCTGAT CCAGTGTATG AAGAAGAAGG TGTGATTCAT TATGGTGTAC CAAATCAACC 900  
 45 AGGAGCAGTC CCAAGAACTT CAACAATGGC ATTAGCACAA GGAAATATTG ATTATATATT 960  
 AGAAATTTGT GACAAAGGCT TAGAACAAGC AATTAAAGAT AATGAAGCCT TAAGTACTGG 1020  
 TGTAACATT TACCAAGGAC AAGTGACAAA TCAAGGATTA GCTTCATCAC ATGACCTAGA 1080  
 50 TTATAAAGAA ATATTAAATG TTATCGAATA GATAGTAATT TAAATGAAAT TGAGTGAAAT 1140  
 GAATATTTTA AATATAGCAT TATAGTTTGG ACTAAAAATT TACAAAACGG AAGGATGTAA 1200

55

## EP 0 786 519 A2

	TCGAAGAAGC TAAAGCAAGC ATTAAACCAT TTATTCGTCG AACACCTCTA ATTAAATCAA	1320
	TGTATTTAAG CCAAAGTATA ACTAAAGGGA ATGTATTTCT AAAATTAGAA AATATGCAAT	1380
5	TCACAGGATC TTTTAAATTT AGAGGCGCTA gCAATnAAAA TTAATCACTT AACAGATGAA	1440
	CAAAAAGAAA AAGGCATTAT CGCAGCATCT GCTGGGgAAC CATGCACAAG GTGTTGCTTT	1500
	AACAGCTAAA TTATTAGGCA TTGATGCAAC GATTGTAATG CCTGAAACAG CACCACAAGC	1560
10	GAAACAACAA GCAACAAAAG GCTATGGGGC AAAGGTTATT TTAAGGTA AAACTTTAA	1620
	CGAAACTAGA CTTTATATGG AAGAATTAGC GAAAGAAAAT GGCATGACAA TCGTTCATCC	1680
	ATATGACGAT AAGTTTGTA TGGCAGGCCA AGGAACAATT GGTTTAGAAA TTTTAGATGA	1740
15	TATTTGGAAT GTGAATACAG TCATCGTACC AGTTGGCGGT GGAGGATTAA TTGCAGGTAT	1800
	TGCCACCGCA TTAAATCAT TTAACCTTC AATTCATATT ATCGGTGTTT AATCTGAGAA	1860
20	TGTTTCATGGT ATGGCTGAGT CTTTCTATAA GAGAGATTTA ACTGAACATC GAGTGGATAG	1920
	CACAATAGCA GATGGTTGTG ATGTAAAAGT TCCTGGTGAA CAAACATATG AAGTAGTTAA	1980
	ACATTTAGTA GATGAATTTA TTCTTGTTAC TGAAGAAGAA ATTGAACATG CTATGAAAGA	2040
25	TTTAATGCAG CGTGCCAAAA TTATTACTGA AGGTGCAGGC GCATTACCAA CAGCTGCAAT	2100
	TTTAAGTGGA AAAATAAACA ATAAATGGCT TGAAGATAAA AATGTTGTTG CATTAGTTTC	2160
	AGGCGGGAAT GTTGACTTAA CTAGAGTTTC AGGTGTCATT GAACATGGAC TGAATATTGC	2220
30	AGATACAAGC AAGGGTGTGG TAGGTTAAAA CATTTAATCT TAAAAATGAG GTGTAATTAT	2280
	GTCAAATGGT AAAGAATTAC AAAAAATAT AGGTTTCTTC TCAGCGTTTG CTATTGTTAT	2340
	GGGACAGTT ATTGGTTCAG GAGTATTCTT TAAATATCA AACGTAACAG AAGTAACAGG	2400
35	AACAGCAGGA ATGGCCTTGT TTGTATGGTT CCTAGGCGGC ATCATTACCA TTTGTGCGGG	2460
	GTTAĀCAGCA GCAGAACTTG CTGCTGCAAT CCCTGAAACA GGTGGCTTAA CGAAGTATAT	2520
40	AGAAATATACA TACGGTGATT TCTGGGGCTT CCTATCAGGT TGGGCGCAAT CATTATTTTA	2580
	TTTTCCAGCT AACGTAGCAG CATTGTCTAT CGTATTTGCG ACACAGCTAA TTAATTTATT	2640
	CCATTTATCT ATAGGTTTCGT TAATACCAAT AGCAATCGCA TCTGCGTTAT CTATTGTGTT	2700
45	GATAAATTC CTAGGTTCAA AAGCAGGCGG AATTTTACAA TCAGTTACTT TAGTAATTAA	2760
	ACTGATTCCA ATCATCGTTA TTGTAATTTT TGGTATTTT CAATCTGGAG ATATCACTTT	2820
	TTCATTAATT CCAACTACAG GTAATTCaGG AAATGGCTTC TTTACAGCAA TTGGTAGTGG	2880
50	TTTATTAGCA ACTATGTTTG CATATGATGG TTGATTTCAT GTAGGAAATG TTGCGGGGGA	2940
	ACTTAAAAAT CCTAAACGCG ATTTACCTTT AGCGATTTCa GTTGGTATCG GTTGTATTAT	3000

55

	TGGTAATTTA AATGCAGCTT CAGATACATC AAAAATATTA TTTGGTGAAA ATGGCGGTAA	3120
	GATTATTACA ATCGGTATAT TAATTTCTGT TTATGGTACG ATCAATGGCT ATACTATGAC	3180
5	TGGTATGCGC GTACCATATG CAATGGCTGA AAGAAAATTA TTGCCATTTA GCCATTTATT	3240
	CGCAAAATTA ACAAATCTG GCGCACCATG GTTTGGCGCA ATTATACAAC TTATAATCGC	3300
	TATCATCATG ATGTCAATGG GAGCATTTGA TACAATTACA AATATGTTAA TCTTTGTTAT	3360
10	TTGGTTGTTC TATTGTATGT CATTTGTTGC GGTAATAATT TTAAGAAAAC GTGAACCAAA	3420
	TATGGAACGA CCATATAAAG TACCGTTATA TCCGATCATA CCTTTAATTG CTATTTTGGC	3480
	AGGATCATTT GTATTAATTA ATACACTGTT TACACAATTT ATATTAGCAA TCATTGGAAT	3540
15	TCTAATAACA GCACTTGGTA TACCAGTTTA TTACTATAAA AAGAAACAAA AAGCAGCATA	3600
	AGGTAAGATA ACTAGCATTG AGAATAAATG GATGGACTAC TAATAAATTT AAAGTTTTAC	3660
	ACATTAAAAT CAAAAACCAT TCAATTATTC TATGGAACAG ACAAATTTCT GTTATGGAAT	3720
20	TTGTCTGTTT TTCAAAAGTA TAGGGAGGCA AATAGAGATG GAAAAGCCGT CAAGAGAGGC	3780
	ATTTGAAGGC AATAATAAGT TGTTAATAGG AATTGTTCTA AGTGTAATAA CGTTTTGGCT	3840
25	ATTTGCACAA TCATTGGTTA ATGTTGTACC AATACTTGAA GATAGTTTCA ATACAGATAT	3900
	TGGAACGGTT AATATCGCCG TTAGTATAAC TGCTTTATTT TCAGGAATGT TTGTAGTAGG	3960
	AGCAGGTGGT CTTGCTGATA AATATGGCAG AATTAACTC ACGAACATTG GTATTATCTT	4020
30	AAATATATTA GGTTCATTAT TAATCATTAT TTCAAATATT CCTTTATTAC TTATTATAGG	4080
	AAGATTAATT CAAGGACTTT CAGCAGCATG TATTATGCCT GCAACTTTGT CTATTATTAA	4140
	GTCATATTAC ATTGGGAAAG ATAGACAACG CGCTTTAAGT TATTGGTCAA TTGGCTCATG	4200
35	GGGCGGCTCT GGTGTTTGTT CATTTTTTGG AGGTGCAGTT GCAACGCTTT TAGGTTGGCG	4260
	TTGGATTTTC ATCCTATCAA TTATAATTTT ATTAATTGCA CTGTTTCTTA TTAAAGGCAC	4320
	ACCTGAACT AAATCTAAAT CGATTCTCTT AAATAAATTT GACATTAAAG GTCTGGTTCT	4380
40	TTTAGTCATT ATGCTCCTCA GTTTAAATAT TTTAATTACT AAAGGATCAG AATTAGGTGT	4440
	AACCTCACTT CTTTTTATTA CTTTATTAGC TATTGCAATT GGATCTTTTA GTTTATTTAT	4500
45	AGTTCTTGAA AAGCGTGCTA CAAATCCTTT AATCGATTTT AAATTATTTA AAAATAAAGC	4560
	TTACACAGGT GCAACAGCTT CAACTTTTTT GTTAAATGGT GTTGCAAGGAA CATTAATAGT	4620
	AGCCAACACA TTTGTTCAAA GAGGTTTAGG ATATTCTTCA TTGCAAGCAG GAAGTTTATC	4680
50	AATCACTTAT TTAGTAATGG TACTAATTAT GATTCGTGTT GGTGAAAAGT TACTTCAAAC	4740
	ACTCGGATGC AAGAAACCAA TGTTAATTGG AACAGGAGTT CTTATTGTCTG GAGAATGTCT	4800

55

	ATTCTTTGGT TTAGGACTAG GGATATATGC TACACCATCA ACAGATACAG CAATTGCAAA	4920
	TGCACCGTTA GAAAAAGTAG GCGTTGCTGC AGGTATCTAT AAAATGGCTT CTGCATTAGG	4980
5	TGGAGCATTI GCGTCGCGAT TGAGTGGTGC AGTATATGCA ATCGTATCAA ATATGaCAAA	5040
	CATTTATACA GGTGcAATGa TTGnCATTAT GGTtaAATGC AGGTATGGGa ATATTATCaT	5100
	TCGTTATCAT TTTGtTACTT GTGcCTAAAC mAAACGACAC TCAATTATGA TAATTGAGAA	5160
10	TTAAATTGAA ATCATACAAG TCGCTACAAT ATTAAACAAA AATATAAACC GATTCTTATG	5220
	TGTCATTATT TTAAATGAAC ATAGGGATTG GTTTTTTATT ACTCTTTTAC GCTACTTTAT	5280
15	TTATAATTAT TATAAATTGT CACAAATTCA ATTTACCTTA CAATATATTT TGTGTTATTA	5340
	TATTCTGGAG CATAAATAAA TTGTTCAACA CATAGTTGTA ATGTGTTTCA ATACTTTTGT	5400
	GATAGATTGC GAAATTGTAT TGAATCGTCA TCGTTTTAAA TTTTAAATG AGAATGGAAT	5460
20	GAGCATTACA ATACACAAGC AATCAAAAGT AAATACATTC ACAACACAAC AGAGACATAA	5520
	CAACAAGATA AGGAGTGAAC AATAGCTGTG AATTATCGTG ATAAAATTCA AAAGTTTAGT	5580
	ATTCGTAAAT ATACAGTTGG TACATTTTCA ACTGTCATTG CGACATTGGT ATTTTtagGA	5640
25	TTCAATACAT CACAAGCACA TGCTGCTGAA ACAAATCAAC CAGCAAGCGT GGTTAAACAG	5700
	AAACAACAAA GTAATAATGA ACAGACTGAG AATCGAGAAT CTCAAGTACA AAATTCTCAA	5760
	AAATCACAAA ATGGTCAATC ATTATCTGCT ACTCATGAAA ATGAGCAACC AAATATTAGT	5820
30	CAAGCTAATT TAGTAGATCA AAAAGTAGCG CAATCATCTA CTACTAATGA TGAACAACCA	5880
	GCATCTCAAA ATGTAAATAC AAAGAAAGAT TCGGCAACGG CTGCGACAAC ACAACCAGAT	5940
	AAAGAACAAA GTAAGCATAA ACAAACGAA AGTCAATCTG CTAATAAAAA TGGAACGAC	6000
35	AATAGAGCGG CTCATGTAGA AAATCATGAA GCAAATGTAG TAACAGCTTC AGATTCATCT	6060
	GATAATTGGTA ACGTACAACA TGACCGAAAT GAATTACAAG CGTTTTTTGA TGCAAATTAT	6120
40	CATGATTATC GCTTTATTGA CCGTGAAAAT GCAGATTCTG GCACATTTAA CTATGTAAAA	6180
	GGCATTTTTG ATAAGATTAA TACGTTATTA GGCAGTAATG ATCCAATAAA CAATAAGAC	6240
	TTGCAACTTG CATACAAAGA ATTGGAACAA GCTGTTGCTT TAATTCGTAC AATGCCTCAA	6300
45	CGTCAACAGA CTAGCCGACG TTCAAATAGA ATTCAAACGC GTTCGGTTGA GTCAAGAGCT	6360
	GCAGAGCCTA GATCAGTATC AGACTATCAA AATGCAAATT CATCATATTA TGTTGAAAAT	6420
	GCTAATGATG GTTCGGGCTA TCCTGTTGGT ACATATATCa ATGCTTCTAG TAAAGGGGCG	6480
50	CCATATAATT TACCAACTAC ACCATGGAAT ACATTGAAGG CCTCTGACTC AAAGGAAATT	6540
	GCTCTTATGA CAGCGAAACA AACTGGAGAC GGGTACCAAT GGGTTATTAA GTTTAATAAA	6600
55		

EP 0 786 519 A2

	GTAGGAAGAA CTGACTTTGT AACAGTTAAT TCAGATGGAA CAAATGTACA ATGGAGTCAT	6720
	GGAGCAGGAG CAGGTGCAAA TAAACCACTT CAACAAATGT GGAATATGG AGTAAATGAT	6780
5	CCTCATCGTT CACATGACTT TAAAATAAGA AATAGAAGTG GCCAAGTAAT ATATGACTGG	6840
	CCAACTGTCC ATATTTATTC TTTAGAAGAT TTATCTAGAG CGAGTGATTA TTTTAGTGAA	6900
	GCTGGAGCGA CACCTGCTAC TAAAGCTTTT GGTAGACAAA ATTTTGAATA TATTAATGGT	6960
10	CAAAAACCTG CTGAATCACC GGGTGTTCCT AAAGTTTATA CTTTCATCGG TCAAGGTGAT	7020
	GCAAGTTATA CAATTTTATT TAAAACACAA GGTCCAACCTG TTAATAAATT GTACTATGCA	7080
	GCAGGTGGGC GTGCTTTAGA GTACAATCAA TTATTTATGT ACAGTCAACT ATACGTCGAA	7140
15	TCAACGCAAG ACCATCAACA ACGTCTTAAT GGTTTAAGAC AAGTGGTTAA TCGTACATAT	7200
	CGCATAGGTA CAACTAAACG TGTAGAAGTG AGTCAAGGAA ATGTACAAAC GAAAAAGGTA	7260
	TTAGAAAGTA CAAACCTAAA TATAGATGAT TTTGTTGATG ATCCTTTAAG TTATGTTAAG	7320
20	ACGCCGAGTA ATAAAGTGTT AGGATTTTAT TCGAATAATG CAAATACTAA TGCTTTTAGA	7380
	CCGGGTGGAG CCAACAATT AAATGAATAT CAATTAAGTC AATTATTTAC TGATCAAAAA	7440
25	TTACAAGAAG CAGCAAGAAC TAGAAACCCA ATAAGATTAA TGATTGGTTT CGACTATCCT	7500
	GATGCTTATG GTAATAGTGA AcTTTAGTTC CTGTTAACTT AACGGTATTA CCTGAAATCC	7560
	AACATAATAT TaAATTCCTT AAAAATGACG ATACTCAAAA TATTGCTGAA AAACCATTTT	7620
30	CAAAACAAGC TGGGCATCCA GTTTTCTATG TATATGCAGG TAACCAAGGG AATGCTTCCG	7680
	TGAATTTAGG TGGTAGCGTA ACATCTATTC AACCATACG TATTAATTTA ACAAGTAATG	7740
	AGAATTTTAC AGATAAAGAT TGGCAAATTA CAGGTATTCC GCGTACATTA CACATTGAAA	7800
35	ACTCGACAAA TAGACCTAAT AATGCCAGAG AACGCAATAT TGAACCTGTT GGTAACCTAT	7860
	TACCGGGGGA TTAACCTTGA ACGATACGTT TTGGACGTAA AGAACAATTA TTCGAAATTC	7920
	GTGTTAAACC ACATACACCA ACAATTACAA CGACAGCTGA GCAATTAAGA GGTACAGCAT	7980
40	TACAAAAGT GCCTGTTAAT ATTTCCGGGA TACCGTTGGA TCCATCGGCA TTGGTTTATT	8040
	TAGTTGCACC AACAAATCAA ACTACGAATG GTGGTAGTGA GGCAGATCAA ATACCATCTG	8100
	GTTATACGAT ACTTGCGACT GGTACACCTG ATGGGGTGCA TAATACAATT ACTATACGAC	8160
45	CGCAAGATTA TGTTGTATTC ATACCACCTG TAGGTAAACA AATTAGAGCA GTAGTTTATT	8220
	ATAATAAGT AGTTGCATCT AATATGAGTA ATGCTGTTAC TATTTTGCCA GATGACATTC	8280
50	CACCAACAAT CAATAATCCT GTTGAATAA ATGCCAATA CTATCGAGGC GACGAACCAA	8340
	CTTTACAATG GGTGTCTCTG ATAGACATTC TGGTATAAAA AATACAATA TTACGACATT	8400

55

## EP 0 786 519 A2

	TACAGGTAGA GTGAGTATGA ATCAGGCATT TAACAGTGAT ATTACATTTA AAGTGTGAGC	8520
	GACAGaCAAT GTCAATAATA CGACAAATGA TAGTCAATCT AAACATGTTT CAATTCATGT	8580
5	AGGTAAAATT ACTGAAGATG CTCATCCGAT TGTATTAGGA AATACTGAGA AAGTTGTAGT	8640
	AGTCAATCCG ACTGCTGTAT CTAATGATGA AAAGCAAAGC ATAATTACTG CCTTTATGAA	8700
	TAAAAACCAA AATATAAGAG GATATTTAGC ATCAACTGAT CCAGTAACTG TCGATAATAA	8760
10	TGGTAATGTC ACATTACATT ACCGTGATGG CTCATCGACA ACGCTTGATG CTACAAATGT	8820
	GATGACATAC GAACCAGTTG TGAACCTGA ATACCAAACCT GTCAATGCTG CTAAAACAGC	8880
15	AACGGTAACG ATTGCTAAAG GACAATCATT TAGTATTGGT GATATTAAAC AATATTTTAC	8940
	TTTAAGTAAT GGACAACCTA TTCCAAGTGG CACATTTACA AATATTACAT CTGATAGAAC	9000
	TATTCCAACT GCACAAGAAG TTAGTCAAAT GAACGCAGGC ACGCAGTTAT ACCATATAAC	9060
20	TGCTACAAAT GCGTATCATA AAGATAGTGA AGACTTCTAT ATTAGTTTGA AAATCATCGA	9120
	TGTGAAACAA CCAGAAGGCG ATCAACGTGT ATATCGTACA TCAACATATG ATTTAACTAC	9180
	TGATGAAATC TCAAAAGTAA AACAAGCATT TATTAATGCA AATAGAGATG TAATTACGCT	9240
25	TGCCGAAGGT GATATTTTCA TTACAAATAC ACCTAATGGT GCTAATGTAA GTACTATTAC	9300
	AGTAAATATT AATAAAGGTC GATTAACGAA ATCATTCGCG TCAAACCTAG CTAATATGAA	9360
	TTTCTTGCGT TGGGTAAATT TCCCACAAGA TTATACAGTG ACATGGACGA ATGCAAAAAT	9420
30	TGCAACAGA CCAACAGATG GTGGTTTATC ATGGTCTGAT GACCATAAAT CTTTAATTTA	9480
	TCGTTATGAT GCTACATTAG GTACTCAAAT TACGACGAAT GATATTTTAA CAATGTTAAA	9540
	AGCAACAACT ACAGTGCCTG GATTGCGAAA TAACATTACT GGTAATGAAA AATCACAAGC	9600
35	AGAAGCTGGC GGAAGACCTA ACTTTAGAAC GACTGGTTAT TCACAATCAA ATGCGACAAC	9660
	TGATGGTCAA CGTCAATTTA CGTTGAATGG TCAAGTGATT CAAGTGTTAG ACATCATCAA	9720
40	CCCTTCAAAC GGTATGGTG GGCAACCTGT TACAAATTCA AATACTCGTG CAAACCATAG	9780
	TAACCAACT GTTGTTAACG TAAACGAACC GGCAGCTAAT GGTGcTGGCG CATTTACAAT	9840
	TGACCACGTT GTAAAAAGTA ATTCTACACA TAATGCAAGT GATGCAGTTT ATAAAGCACA	9900
45	GTTATACTTA ACGCCATATG GTCCAAAACA ATATGTTGAA CATTTAAATC AAAATACAGG	9960
	AAATACTACT GACGCTATTA ACATTTATTT TGTACCAAGT GACTTAGTGA ATCCAACAAT	10020
	TTCAGTAGGT AATTACACTA ATCATCAAGT GTTCTCAGGT GAAACATTTA CAAATACTAT	10080
50	TACAGCGAAT GATAACTTTG GTGTGCAATC TGTAAGTGA CCAAATACAT CACAAATTAC	10140
	AGGTACTGTT GATAATAACC ATCAACATGT TTCTGCAACG GCACCAAATG TGACATCAGC	10200

55



EP 0 786 519 A2

	GTTCAATGTA ACAGTGAAAC CTTTGCGTGA TAAATATCGA GTTGGTACTT CATCAACGGC	10320
	TGCTAATCCT GTGAGAATTG CCAATATTTT GAATAATGCG ACAGTATCAC AAGCTGATCA	10380
5	AACGACAATT ATTAATTCGT TAACGTTTAC TGAAACAGTA CCAAATAGAA GTTATGCAAG	10440
	AGCAAGTGCG AATGAAATCA CTAGTAAAC AGTTAGTAAT GTCAGTCGTA CTGGAAATAA	10500
	TGCCAATGTg CACAGTAACT GTTACTTATC AAGATGGAAC AACATCAACA GTGACTGTAC	10560
10	CTGTAAAGCA TGTCATTCCA GAAATCGTTG CACATTGCGA TTACTGTGTA CAAGGCCAAG	10620
	ACTTCCCAGC AGGTAATGGT TCTAGTGCAT CAGATTACTT TAAGTTATCT AATGGTAGTG	10680
	ACATTGCAGA TGCAACTATT ACATGGGTAA GTGGACAAGC GCCAAATAAA GATAATACAC	10740
15	GTATTGGTGA AGATATAACT GTAACGCAC ATATCTTAAT TGATGGCGAA ACAACGCCGA	10800
	TTACGAAAAC AGCAACATAT AAAGTAGTAA GAACTGTACC GAAACATGTC TTTGAAACAG	10860
20	CCAGAGGTGT TTTATACCCA GGTGTTTCAG ATATGTATGA TGCGAAACAA TATGTTAAGC	10920
	CAGTAAATAA TTCTTGGTCG ACAAATGCGC AACATATGAA TTTCCAATTT GTTGAACAT	10980
	ATGGTCCTAA CAAAGATGTT GTAGGCATAT CTAAGTCTCT TATTAGAGTG ACATATGATA	11040
25	ATAGACAAAC AGAAGATTTA ACTATTTTAT CTAAAGTTAA ACCTGACCCA CCTAGAATTG	11100
	ACGCAAATC TGACATAT AAAGCAGGTC TTACAAACCA AGAAATTAAA GTTAATAACG	11160
	TATTAAATAA CTCGTCAGTA AAATTATTTA AAGCAGATAA TACACCATTA AATGTCACAA	11220
30	ATATTACTCA TGGTAGCGGT TTTAGTTCGG TTGTGACAGT AAGTGACGCG TTACCAAATG	11280
	GCGGAATTAA AGCAAAATCT TCAATTTCAA TGAACAATGT GACGTATACG ACGCAAGACG	11340
	AACATGGTCA AGTTGTTACA GTAACAAGAA ATGAATCTGT TGATTCAAAT GACAGTGCAa	11400
35	CAGTAACAGT GACACCACAA TTACAAGCAA CTAAGTGAAG CGCTGTATTT ATTAAAGGTG	11460
	GCGAAGGTTT TGATTTTCGGA CACGTAGAAA GATTTATTCA AAACCCGCCA CATGGGGCAA	11520
40	CGGTTCATG GCATGATAGT CCAGATACAT GGAAGAATAC AGTCGGTAAC ACTCATAAAA	11580
	CTGCGGTTGT AACATTACCT AATGGTCAAG GTACGCGTAA TGTTGAAGTT CCAGTCAAAG	11640
	TTTATCCAGT TGCTAATGCA AAGGCGCCAT CACGTGATGT GAAAGGTCAA AATTTGACTA	11700
45	ATGGAACGGA TGCGATGAAC TACATTACAT TTGATCCAAA TACAAACACA AATGGTATCA	11760
	CTGCAGCATG GGCAATAGA CAACAACCAA ATAACCAACA AGCAGGCGTG CAACATTTAA	11820
	ATGTCGATGT CACATATCCA GGTATTTTCA CTGCTAAACG AGTTCCTGTT ACTGTTAATG	11880
50	TATATCAATT TGAATTCCTT CAACTACTT ATACGACAAC GGTTGGAGGC ACTTTAGCAA	11940
	GTGGTACGCA AGCATCAGGA TATGCACATA TGCAAAATGC TACTGGTTTA CCAACAGATG	12000

55

TGAATAAACC GAATGTGGCT AAAGTCGTTA ACGCAAAATA TGACGTCATC TATAACGGAC 12120  
 ATACTTTTGC AACATCTTTA CCAGCGAAAT TTGTAGTAA AGATGTGCAA CCAGCGAAAC 12180  
 5 CAACTGTGAC TGAAACAGCG GCAGGAGCGA TTACAATTGC ACCTGGAGCA AACCACAG 12240  
 TGAATACACA TGCCGGTAAC GTAACGACAT ACGCTGATAA ATTAGTTATT AAACGTAATG 12300  
 GTAACGTTGT GACGACATTT ACACGTCGCA ATAATACGAG TCCATGGGTG AAAGAAGCAT 12360  
 10 CTGCAGCAAC TGTAGCAGGT ATTGCTGGAA CTAATAATGG TATTACTGTT GCAGCAGGTA 12420  
 CTTTCAACCC TGCTGATACA ATTCAAGTTG TTGCAACGCA AGGAAGCGGA GAGACAGTGA 12480  
 GTGATGAGCA ACGTAGTGAT GATTTACACAG TTGTCGCACC ACAACCGAAC CAAGCGACTA 12540  
 15 CTAAGATTG GCAAAATGGT CATATTGATA TCACGCCTAA TAATCCATCA GGACATTTAA 12600  
 TTAATCCAAC TCAAGCAATG GATATTGCTT ACGTGAATA AGTGGGTAAT GGTGCAGAAC 12660  
 ATAGTAAGAC AATTAATGTT GTTCGTGGTC AAAATAATCA ATGGACAATT GCGAATAAGC 12720  
 20 CTGACTATGT AACGTTAGAT GCACAACTG GTAAAGTGAC GTTCAATGCC AATACTATAA 12780  
 AACCACATTC ATCAATCACA ATTACTCCGA AAGCAGGTAC AGGTCACTCA GTAAGTAGTA 12840  
 ATCCAAGTAC ATTAAGTACA CCGGCAGCTC ATACTGTCAA CACAAGTAA ATTGTGAAAG 12900  
 ATTATGGTTC AAATGTAACA GCAGCTGAAA TTAACAATGC AGTTCaAGTT GCTAATAAAC 12960  
 GTAGTGAAC GATTAAAAAT GGCACAGCAA TGCCTACTAA TTTAGCTGGT GGTAGCACAA 13020  
 30 CGACGATTCC TGTGACAGTA ACTTACAATG ATGGTAGTAC TGAAGAAGTA CAAGAGTCCA 13080  
 TTTTCACAAA AGCGGATAAA CGTGAGTTAA TCACAGCTAA AAATCATTTA GATGATCCAG 13140  
 TAAGCACTGA AGGTAAAAAG CCAGGTACAA TTACGCAGTA CAATAATGCA ATGCATAATG 13200  
 35 CGCAACAACA AATCAATACT GCGAAAACAG AAGCACAACA AGTGATTAAT AATGAGCGTG 13260  
 CAACACCACA ACAAGTTTCT GACGCACTAA CTAAAGTTCTG TGCAGCACAA ACTAAGATTG 13320  
 ATCAAGCTAA AGCATTACTT CAAAATAAAG AAGATAATAG CCAATTAGTA ACGTCTAAAA 13380  
 40 ATAACCTACA AAGTCTGTG AACCAAGTAC CATCACTGC TGGTATGACG CAACAAAGTA 13440  
 TTGATAACTA TAATGCGAAG AAGCGTGAAG CAGAACTGA AATAACTGCA GCTCAACGTG 13500  
 TTATTGACAA TGGCGATGCA ACTGCACAAC AAATTTTCTG TGAAAAACAT CGTGTCGATA 13560  
 ACGCATTAAC AGCATTAAC CAAGCGAAAC ATGATTTAAC TGCAGATACA CATGCCTTAG 13620  
 AGCAAGCAGT GCAACAATTG AATCGCACAG GTACAACGAC TGGTAAGAAG CCGGCAAGTA 13680  
 50 TTAAGTCTTA CAATAATTCG ATTCGTGCAC TTCAAAGTGA CTTAACAAGT GCTAAAAATA 13740  
 GCGCTAATGC TATTATTCAA AAGCCAATAA GAACAGTACA AGAAGTGCAA TCTGCGTTAA 13800

55

CTGATAATAG TGCTTTAAAA ACTGCTAAGA CGAAACTTGA TGAAGAAATC AATAAATCAG 13920  
 TAACTACTGA TGGTATGACA CAATCATCAA TCCAAGCATA TGAAAATGCT AAACGTGCGG 13980  
 5 GTCAAACAGA ATCAACAAAT GCACAAAATG TTATTAACAA TGGTGATGCG ACTGACCAAC 14040  
 AAATTGCCGC AGAAAAACA AAAGTAGAAG AAAAATATAA TAGCTTAAAA CAAGCAATTG 14100  
 CTGGATTAACT TCCAGACTTG GCACCATTAC AAAGTGCATA AACTCAGTTG CAAAATGATA 14160  
 10 TTGATCAGCC AACGAGTACG ACTGGTATGA CAAGCGCATC TATTGCAGCA TTTAATGAAA 14220  
 AACTTTCAGC AGCTAGAACT AAAATTCAAG AAATTGATCG TGTATTAGCC TCACATCCAG 14280  
 ATGTTGCGAC AATACGTCAA AACGTGACAG CAGCGAATGC CGCTAAATCA GCACTTGATC 14340  
 15 AAGCACGTAA TGGCTTAACT GTCGATAAAG CGCCTTTAGA AAATGCGAAA AATCAACTAC 14400  
 AACATAGTAT TGACACGCAA ACAAGTACAA CTGGTATGAC ACAAGACTCT ATAAATGCAT 14460  
 20 ACAATGCGAA GTTAACAGCT GCACGTAATA AGATTCAACA AATCAATCAA GTATTAGCAG 14520  
 GTTCACCGAC TGTAGAACAA ATTAATACAA ATACGTCTAC AGCAAATCAA GCTAAATCTG 14580  
 ATTTAGATCA TGCACGTCAA GCTTTAACAC CAGATAAAGC GCCGCTTCAA ACTGCGAAAA 14640  
 25 CGCAATTAGA ACAAAGCATT AATCAACCAA CGGATACAAC AGGTATGACG ACCGCTTCGT 14700  
 TAAATGCGTA CAACCAAAAA TTACAAGCAG CGCGTCAAAA GTTAACTGAA ATTAATCAAG 14760  
 TGTTGAATGG CAACCCAACT GTCCAAAATA TCAATGATAA AGTGACAGAG GCAAACCAAG 14820  
 30 CTAAGGATCA ATTAATACA GCACGTCAAG GTTTAACATT AGATAGACAG CCAGCGTTAA 14880  
 CAACATTACA TGGTGCATCT AACTTAAACC AAGCACAACA AAATAATTTC ACGCAACAAA 14940  
 TTAATGCTGC TCAAAATCAT GctGCGCTTG AAACAATTAA GTCTAACATT ACGGCTTTAA 15000  
 35 ATACTGCGAT GACGAAATTA AAAGACAGTG TTGCGGATAA TAATACAATT AAATCAGATC 15060  
 AAAATTACAC TGACGCAACA CCAGCTAATA AACAAGCGTA TGATAATGCA GTTAATGCGG 15120  
 40 CTAAAGGTGT CATTGGAGAA ACGACTAATC CAACGATGGA TGTTAACACA GTGAACCAAA 15180  
 AAGCAGCATC TGTTAAATCG ACGAAAGATG CTTTAGATGG TCAACAAAAC TTACAACGTG 15240  
 CGAAAACAGA AGCAACAAAT GCGATTACGC ATGCAAGTGA TTTAAACCAA GCACAAAAGA 15300  
 45 ATGCATTAACT ACAACAAGTG AATAGTGcAC AAAACGTGCA AGCAGTAAAT GATATTAAAC 15360  
 AAACGACTCA AAGCTTAAAT ACTGCTATGA CAGGTTTAAA ACGTGGCGTT GCTAATCATA 15420  
 ACCAAGTCGT ACAAAGTGAT AATTATGTCA ACGCAGATAC TAATAAGAAA AATGATTACA 15480  
 50 ACAATGCATA CAACCATGCG AATGACATTA TTAATGGTAA TGCACAACAT CCAGTTATAA 15540  
 CACCAAGTGA TGTTAACAAT GCTTTATCAA ATGTCACAAG TAAAGAACAT GCATTGAATG 15600

55

	ATTAAATAA TGCACAACGT CAAAACCTTAC AATCGCAAAT TAATGGTGCG CATCAAATTG	15720
	ATGCAGTTAA TACAATTAAG CAAAATGCAA CAACTTGAA TAGTGCAATG GGTAACCTAA	15780
5	GACAAGCTGT TGCAGATAAA GATCAAGTGA AACGTACAGA AGATTATGCG GATGCAGATA	15840
	CAGCTAAACA AAATGCATAT AACAGTGCAG TTTCAAGTGC CGAAACAATC ATTAATCAAA	15900
	CAACAAATCC AACGATGTCT GTTGATGATG TTAATCGTGC AACTTCAGCT GTTACTTCTA	15960
10	ATAAAATGC ATTAAATGGT TATGAAAAAT TAGCACAATC TAAACAGAT GCTGCAAGAG	16020
	CAATTGATGC ATTACCACAT TTAAATAATG CACAAAAAGC AGATGTTAAA TCTAAAATTA	16080
	ATGCTGCATC AAATATTGCT GCGGTAAATA CTGTTAAACA ACAAGGTACA GATTTAAATA	16140
15	CAKCGATGGg TAACCTGCAA GGTGCAATCA ATGATGAACA AACGACGCTT AATAGTCAAA	16200
	ACTATCAAGA TCGCACACCT AGTAAGAAAA CAGCATACAC AAATGCGGTA CAAGCTGCGA	16260
20	AAGATATTTT AAATAAATCA AATGGTCAAA ATAAACGAA AGATCAAGTT ACTGAAGCGA	16320
	TGAATCAAGT GAATTCTGCT AAAAATAACT TAGATGGTAC GCGTTTATTA GATCAAGCGA	16380
	nCAAaCAGCA AAACAGCAGT TAAATAATAT GACGCATTTA ACAACTGCAC AAAAAACGAA	16440
25	TTTAACAAAC CAAATTAATA GTGGTACTAC TGTCGCTGGT GTTCAAACGG TTCAATCAAA	16500
	TGCCAATACA TTAGATCAAG CCATGAATAC GTTAAGACAA AGTATTGCCA ACAAAGATGC	16560
	GACTIONAGCA AGTGAAGATT ACGTAGATGC TAATAATGAT AAGCAAACAG CATATAACAA	16620
30	CGCAGTAGCT GCTGCTGAAA CGATTATTAA TGCTAATAGT AATCCAGAAA TGAATCCAAG	16680
	TACGATTACA CAAAAGCAG AGCAAGTGAA TAGTTCTAAA ACGGCACTTA ACGGTGATGA	16740
	AACTTAGCT GCTGCAAAAC AAAATGCGAA AACGTACTTA AACACATTGA CAAGTATTAC	16800
35	AGATGCTCAA AAGAACAATT TGATTAGTCA AATTACTAGT GCGACAAGAG TGAGTGGTGT	16860
	TGATACTGTA AAACAAAATG CGCAACATCT AGACCAAGCT ATGGCTAGCT TACAGAATGG	16920
40	TATTAACAAC GAATCTCAAG TGAAATCATC TGAGAAATAT CGTGATGCTG ATACAAATAA	16980
	ACAACAAGAG TATGATAATG CTATTACTGC AGCGAAAGCG ATTTTAAATA AATCGACAGG	17040
	TCCAAACACT GCGCAAAATG CAGTTGAAGC AGCATTACAA CGTGTTAATA ATGCGAAAGA	17100
45	TGCATTGAAT GGTGATGCAA AATTAATTGC AGCTCAAAAC GCAGCGAAAC AACATTTAGG	17160
	TACTTTAACG CATATCACTA CAGCTCAACG TAATGATTTA ACAAATCAAA TTTACAAGC	17220
	TACAACTTA GCTGGTGTG AATCTGTTAA ACAAATGCG AATAGTTTAG ATGGTGCTAT	17280
50	GGGTAACCTA CAAACGGCTA TCAACGATAA GTCAGGAACA TTAGCGAGCC AAAACTTCTT	17340
	GGATGCTGAT GAGCAAAAAC GTAATGCATA CAATCAAGCT GTATCAGCAG CCGAAACCAT	17400

55

EP 0 786 519 A2

	TGTTAATAAT GCGAAACATG CATTAAATGG TACGCAAAAC TTAAACAATG CGAAACAAGC	17520
	AGCGATTACA GCAATCAATG GCGCATCTGA TTTAAATCAA AAACAAAAAG ATGCATTAAA	17580
5	AGCACAAGCT AATGGTGCTC AACGCGTATC TAATGCACAA GATGTACAGC ACAATGCGAC	17640
	TGAACTGAAC ACGGCAATGG GCACATTAAA ACATGCCATC GCAGATAAGA CGAATACGTT	17700
	AGCAAGCAGT AAATATGTTA ATGCCGATAG CACTAAACAA AATGCTTACA CAACTAAAGT	17760
10	TACCAATGCT GAACATATTA TTAGCGGTAC GCCAACGGTT GTTACGACAC CTTCAGAAGT	17820
	AACAGCTGCA GCTAATCAAG TAAACAGCGC GAAACAAGAA TTAAATGGTG ACGAAAGATT	17880
	ACGTGAAGCA AAACAAAACG CCAATACTGC TATTGATGCA TTAACACAAT TAAATACACC	17940
15	TCAAAAAGCT AAATTTAAAG AACAAAGTGGG ACAAGCCAAT AGATTAGAAG ACGTACAAAC	18000
	TGTTCAAACA AATGGACAAG CATTGAACAA TGCAATGAAA GGCTTAAGAG ATAGTATTGC	18060
	TAACGAAACA ACAGTCAAAA CAAGTCAAAA CTATACAGAC GCAAGTCCGA ATAACCAATC	18120
20	AACATATAAT AGCGCTGTGT CAAATGCGAA AGGTATCATT AATCAAATA ACAATCCGAC	18180
	TATGGATACT AGTGCGATTA CCCAAGCTAC AACACAAGTG AATAATGCTA AAAATGGTTT	18240
	AAACGGTGCT GAAAACCTAA GAAATGCACA AAACACTGCT AAGCAAACT TAAATACATT	18300
25	ATCACACTTA ACAAATAACC AAAAATCTGC CATCTCATCA CAAATTGATC GTGCAGGTCA	18360
	TGTGAGTGAG GTAACCTGCTA CTAAAAATGC AGCAACTGAG TTGAATACGC AAATGGGTAA	18420
	CTTGGAACAA GCTATCCATG ATCAAAACAC AGTTAAACAA AGTGTTAAAT TTACTGATGC	18480
	AGATAAAGCT AAACGTGATG CGTATACAAA TGCGGTAAGC AGAGCTGAAG CAATTCTGAA	18540
	TAAACGCAA GGTGCAAATA CGTCTAAACA AGATGTTGAA GCGGCTATTC AAAATGTTTC	18600
35	AAGTGCTAAA AATGCATTGA ATGGTGATCA AAACGTTACA AATGCGAAGA ATGCAGCTAA	18660
	AAATGSCATTA AATAACTTAA CGTCAATTAA TAATGCACAA AAACGTGACT TAACAATAA	18720
	AATTGATCAA GCAACAACCTG TAGCTGGTGT TGAAGCTGTA TCTAATACGA GTACACAATT	18780
40	GAATACAGCG ATGGCTAACT TGCAAAATGG TATTAATGAT AAAACAAATA CACTAGCAAG	18840
	TGAAAACCTAT CATGATGCTG ATTCAGATAA GAAAACCTGCT TATACTCAAG CCGTTACGAA	18900
	CGCAGAAAAT ATTTTAAATA AAAATAGTGG ATCAAATTTA GACAAAACCTG CCGTTGAAAA	18960
45	CGCGTTGTCA CAAGTTGCTA ATGCGAAAGG TGCCCTAAAT GGTAACCATA ATTTAGAGCA	19020
	AGCTAAATCA AATGCAAACA CTACTATAAA CGGACTTCAA CATTTAACAA CTGCTCAAAA	19080
	AGATAAATTG AAACAACAAG TGCAACAAGC ACAAATGTT GCAGGTGTAG ATACTGTTAA	19140
50	ATCAAGTGCC AACACATTAA ATGGTGCTAT GGGTACGTTA AGAAATAGCA TACAAGATAA	19200

55

	TAACAATGCT GTTGATAGTG CTAATGGTGT CATTAAATGCA ACAAGCAATC CAAATATGGA	19320
	TGCTAATGCA ATTAACCAAA TCGCTACACA AGTGACATCA ACGAAAAATG CATTAGATGG	19380
5	TACACATAAT TTAACGCAAG CGAAACAAAC AGCAACAAAT GCCATCGATG GTGCTACTAA	19440
	CTTAAATAAA GCGCAAAAAG ATGCGTTAAA AGCACAAAGTT ACAAGTGCGC AACGTGTTGC	19500
	AAATGTAACA AGTATCCAAC AAAGTCAAAA TGAAGTTAAT ACAGCTATGG GTCAATTACA	19560
10	ACATGGTATT GATGATGAAA ATGCAACAAA ACAAACTCAA AAATATCGTG ACGcTGAACA	19620
	AAGTAAGAAA ACTGCTTATG ATCAAGCTGT AGCTGCTGCG AAAGCAATTT TAAATAAACA	19680
	AACAGGTTCA AATTCAGATA AAGCAGCAGT TGACCGTGCA TTACAACAAG TAACAAGTAC	19740
15	GAAAGATGCA TTGAATGGTG ATGCAAAACT GGCAGAAGCG AAAGCGGCAG CTAAACAAAA	19800
	CTTAGGCACT TTAAACCATA TTACGAATGC ACAACGTACT GACTTAGAAG GCCAAATCAA	19860
20	TCAAGCGACG ACTGTTGATG GCGTTAATAC TGTA AAAACA AATGCCAATA CATTAGACGG	19920
	CGCAATGAAT AGCTTACAAG GTTCAATCAA TGATAAAGAT GCGACATTAA GAAATCAAAA	19980
	TTATCTTGAT GCGGATGAAT CAAAACGAAA TGCATATACG CAAGCTGTCA CAGCGGCTGA	20040
25	AGGCATTTTA AATAAACAAA CTGGTGGTAA CACATCTAAA GCAGACGTTG ATAATGCATT	20100
	AAATGCAGTT ACAAGAGCGA AAGcGgCTTT AAATGGTGCT GACAACTTAA GAAATGCGAA	20160
	AACTTCAGCA ACAAATACGA TTGATGGTTT ACCTAACTTA ACACAATTAC AAAAAGACAA	20220
30	CTTGAAGCAT CAAGTTGaaC AAGCGCAAAA TGTAGCAGGT GTAAATGGTG TTAAAGATAA	20280
	AGGTAATACG TTAAATACTG CCATGGGTGC ATTACGTACA AGTATCCAAA ATGATAATAC	20340
	GACGAAAACA AGTCAAAATT ATCTTGATGC ATCTGACAGC AACAAAAATA ATTACAATAC	20400
35	TGCTGTAAAT AATGCAAATG GTGTTATTAA TGCAACGAAC AATCCAAATA TGGATGCTAA	20460
	TGCGATTAAAT GGCATGGCAA ATCAAGTCAA TACAACAAAA GCAGCGTTAA ATGGTGACAA	20520
	AAACTTAGCT CAAGCTAAAA CAAATGCGAC GAACACAATT AACAAACGCAC ATGACTTAAA	20580
40	CCAAAAACAA AAAGATGCAT TAAAAACACA AGTTAACAAT GCACAACGTG TATcTGATGC	20640
	AAATAACGTT CAACACACTG CAACTGAATT GAACAGTGCG ATGACAGCAC TTAAAGCAGC	20700
45	TATTGCTGAT AAAGAAAGAA CAAAAGCAAG CGGTAATTAT GTCAATGCTG ATCAAGAAAA	20760
	ACGTCAAGCG TATGATTCAA AAGTGAATAA CGCTGAAAAT ATCATTAGTG GTACACCGAA	20820
	TGCGACATTA ACAGTCAATG ACGTAAATAG TGCGGCATCA CAAGTCAATG CGGCTAAAAC	20880
50	AGCATTAAAT GGTGATAACA ACTTACGTGT AGCGAAAGAG CATGCCAACA ATACAATTGA	20940
	CGGCTTAGCA CAATTGAATA ATGCACAAAA AGCAAAATTA AAAGAACAAG TTCAAAGTGC	21000

55

EP 0 786 519 A2

	GAAAGGCTTA AGAGATAGTA TTGCGAATGA AGCAACAATT AAAGCAGGTC AAAACTACAC	21120
	TGACGCAAGT CCAAATAATC GTAACGAGTA CGACAGTGCA GTTACTGCAG CAAAAGCAAT	21180
5	CATTAATCAA ACATCGAACC CAACGATGGA ACCAAATACT ATTACGCAAG TAACATCACA	21240
	AGTGACAACCT AAAGAACAGG CATTAAATGG TGCGCGAAAC TTAGCTCAAG CTAAGACAAC	21300
	TGCGAAAAAC AACTTGAATA ACTTAACATC AATTAACAAT GCACAAAAAG ATGCGTTAAC	21360
10	GCGTAGcATT GATGGTGCAA CAACAGTAGC TGGTGTAAT CAAGAACTG CAAAAGCAAC	21420
	AGAATTAAAT AACGCAATGC ATAGTTTACA AAATGGTATC AATGATGAGA CACAAACAAA	21480
	ACAACTCAG AAATACCTAG ATGCAGAGCC AAGTAAGAAA TCAGCTTATG ATCAAGCAGT	21540
15	AAATGCAGCG AAAGCAATTT TAACAAAAGC TAGTGGTCAA AATGTAGACA AAGCAGCAGT	21600
	TGAACAAGCA TTGCAAAATG TGAACAGTAC GAAGACGGCG TTGAACGGTG ATGCGAAATT	21660
20	AAATGAAGCT AAAGCAGCTG CGAAACAAAC GTTAGGTACA TTAACACACA TTAATAATGC	21720
	ACAACGTACA GCGTTAGACA ATGAAATTAC ACAAGCAACA AATGTTGAAG GTGTTAATAC	21780
	AGTTAAAGCC AAAGCGCAAC AATTAGATGG TGCTATGGGT CAATTAGAAA CATCAATTCG	21840
25	TGATAAAGAC ACGACGTTAC AAAGTCAAAA TTATCAAGAT GCTGATGATG CTAAACGAAC	21900
	TGCTTATTCT CAAGCAGTAA ATGCAGCAGC AACTATTTTA AATAAAACag CTGGCGGTAA	21960
	TACACCTAAA GCAGATGTTG AAAGAGCAAT GCAAGCTGTT ACACAAGCAA ATACTGcATT	22020
30	AAACGGTATT CAmAACTTAG ATCGTGCGAA ACaGCTGCT AACACAGCGA TTACAAATGC	22080
	TTCGGACTTA AATACAAAAC mAAAAGAAGC ATTAaAGCA CAAGTAACAA GTGCAGGACG	22140
	TGTATCTGCA GCAAATGGTG TTGAACATAC TGCGACTGAA TTAAATACTG CGATGACAGC	22200
35	TTTAAAGCGT GCCATTGCTG ATAAAGCTGA GACAAAAGCT AGTGGTAACT ATGTCAATGC	22260
	TGATGCGAAT AAACGTCAAG CATATGATGA AAAAGTTACA GCTGCCGAAA ATATCGTTAG	22320
40	TGGTACACCA ACACCAACGT TAACACCAGC AGATGTTACA AATGCAGCAA CGCAAGTAAC	22380
	GAATGCTAAG ACGCAGTTAA ACGGTAATCA TAATTTAGAA GTAGCGAAAC AAAATGCTAA	22440
	CACTGCAATT GATGGTTTAA CTTCTTTAAA TGGTCCGCAA AAAGCAAAAC TTAAAGAACA	22500
45	AGTGGGTCAA GCGACGACGT TGCCAAATGT TCAAAGTGT CGTGATAATG CACAAACATT	22560
	AAACACTGCA ATGAAAGGTC TACGAGATAG CATTGCGAAT GAAGCAACGA TTAAAGCAGG	22620
	TCAAACTAC ACAGATGCAA GTCAAAACAA ACAAACTGAC TACAACAGTG CAGTCACTGC	22680
50	AGCAAAAGCA ATCATTGGTC AAACAAC TAGTCAATG AATGCGCAAG AAATTAATCA	22740
	AGCGAAAGAC CAAGTGACAG CTAAACAACA AGCGTTAAAC GGTCAAGAAA ACTTAAGAAC	22800

55

	AGATGCAGTG AAACGTCAAA TCGAAGGTGC AACGCATGTT AATGAAGTAA CACAAGCACA	22920
	AAATAATGCG GATGCaTTAA ATACAGCTAT GACGAACTTG AAAAATGGTA TTCAAGATCA	22980
5	GAATACGATT AAGCAAGGTG TTAACCTTCAC TGATGCCGAC GAAGCGAAAC GTAATGCATA	23040
	TACAAATGCA GTGACGCAAG CTGAACAAAT TTTAAATAAA GCACAAGGTC CAAATACTTC	23100
	AAAAGACGGT GTCGAACTG CGTTAGaAA TGTACAACGT GCTAAAAACG AATTGAACGG	23160
10	TAATCAAAAT GTTGCGAAG CTAAGACAAC TGCGAAAAAT GCATTGAATA ACCTAACATC	23220
	AATTAATAAT GCACAAAAAG AAGCATTGAA ATCACAAATT GAAGGTGCGA CAACAGTTGC	23280
15	AGGTGTAAAT CAAGTGCTA CAACGGCATC TGAATTAAAT ACAGCAATGA GCAACTTACA	23340
	AAATGGTATT AATGATGAAG CAGCTACAAA AGCAGCGCTT AATGGTACTC AAAACCTTGA	23400
	AAAAGCTAAA CAACACGCAA ATACAGCAAT TGACGGTTTA AGCCATTTAA CAAATGCACA	23460
20	AAAAGAGGCA TTAAACAAT TGGTACAACA ATCGACTACT GTTGCGAAG CACAAGGTAA	23520
	TGAGCAAAAA GCAACAATG TTGATGCAGC AATGGACAAA TTACGTCAAA GTATTGCAGA	23580
	TAATGCGACA ACAAACAAA ACCAAAATTA TACTGATGCA AGTCAGAATA AAAAGGATGC	23640
25	GTACAATAAT GCTGTCACAA CTGCACAAGG TATTATTGAT CAAACTACAA GTCCAACTTT	23700
	AGATCCGACT GTTATCAATC AAGCTGCTGG ACAAGTAAGC ACAACTAAAA ATGCATTAAA	23760
	TGGTAATGAA AACCTAGAGG CAGCGAAACA ACAAGCGTCA CAATCATTAG GTTCATTAGA	23820
30	TAACTTAAAT AATGCGCAAA AACAAACAGT TACTGATCAA ATTAATGGCG CGCATACTGT	23880
	TGATGAAGCA AATCAAATTA AGCAAAATGC GCAAACTTA AATACAGCGA TGGGTAACCTT	23940
	GAAACAAGCG ATAGcTGACA AAGATGCTAC GAAAGCGACA GTTAACTTCA CTGATGCAGA	24000
35	TCAAGCAAAA CAACAAGCAT ATAACaCTGC TGTTACAAAT GCTGAAAATA TCATTTCAAA	24060
	AGCTAATGGC GGCAATGCAA CACAAGCTGA AGTTGAACAA GCAATCAAAC AAGTTAATGC	24120
40	TGCAAAACAA GCATTAAATG GTAATGCCAA CGTTCAACAT GCAAAAGACG AAGCAACAGC	24180
	ATTAATTAAT AGCTCTAATG ACCTTAACCA AGCACAAAAA GACGCATTAA AACAACAAGT	24240
	TCAAAATGCA ACTACTGTAG CTGGTGTAAG CAATGTTAAA CAAACAGCAC AAGAGTTAAA	24300
45	CAATGCTATG ACACAATTAA AACAAGGCAT TGCAGATAAA GAACAAACAA AAGCTGATGG	24360
	TAACTTTGTC AATGCAGATC CTGATAAGCA AAATGCATAT AATCAAGCAG TAGCGAAAGC	24420
	TGAAGCATTa ATTAGTGcTA CGCCTGATGT TGTCGTTACA CCTAGCGAAA TTACTGCAGC	24480
50	GTTAAATAAA GTTACGCAAG CTAAAAATGA TTTAAATGGT AATACAACT TAGCAACGGC	24540
	GAAACAAAAT GTTCAACATG CTATTGATCA ATTGCCAAAC TTAAACCAAG CGCAACGTGA	24600

55



EP 0 786 519 A2

	AGCGGCGACA	ACGCTTAATG	ACGCGATGAC	ACAATTGAAA	CAAGGTATTG	CGAATAAAGC	24720
	ACAAATTAAA	GGTAGCGAGA	ACTATCACGA	TGCTGATACT	GACAAGCAAA	CAGCATATGA	24780
5	TAATGCAGTA	ACAAAAGCAG	AAGAATTGTT	AAAACAAACA	ACAAATCCAA	CAATGGATCC	24840
	AAATACAATT	CAACAAGCAT	TAAGTAAAGT	GAATGACACA	AATCAAGCAC	TTAACGGTAA	24900
	TCAAAAATTA	GCTGATGCCA	AACAAGATGC	TAAGACAACA	CTTGGTACAC	TAGATCATTT	24960
10	AAATGATGCT	CAAAAACAAG	CGCTAACCAAC	TCAAGTTGAA	CAAGCACCAG	ATATTGCAAC	25020
	AGTTAATAAT	GTTAAGCAAA	ATGCTCAAAA	TCTGAATAAT	GCTATGACTA	ACTTAAACAA	25080
	TGCATTACAA	GATAAACTG	AGACATTAAA	TAGCATTAAAC	TTTACTGATG	CAGATCAAGC	25140
15	TAAGAAAGAT	GCTTATACTA	ATGCGGTTTC	ACATGCAGAA	GGTATTTTAT	CTAAAGCAAA	25200
	TGGCAGCAAT	GCAAGTCAAA	CTGAAGTGA	ACAAGCGATG	CAACGTGTGA	ACGAAGCGAA	25260
20	ACAAGCATTG	AATGGTAATG	ACAATGTACA	ACGTGCAAAA	GATGCAGCGA	AACAAGTGAT	25320
	TACAAATGCA	AATGATTTAA	ATCAAGCAAT	GACACAATTG	AAACAAGGTA	TTGCAGATAA	25380
	AGACCAAAC	AAAGCAAATG	GTAACCTTGT	CAATGCTGAT	ACTGATAAGC	AAAATGCTTA	25440
25	CAACAATGCG	GTAGCACATG	CTGAACAAAT	AATTAGTGGT	ACACCAAATG	CAAACGTGGA	25500
	TCCACAACAA	GTGGCTCAAG	CGTTACAACA	AGTGAATCaA	GCTAAGGGTG	ATTTAAACGG	25560
	TAACCATAAC	TTACAAGTTG	CTAAAGACAA	TGCAAATACA	GCCATTGATC	AGTTACCAAA	25620
30	CTTAAATCAA	CCACAAAAAA	CAGCATTAAA	AGACCAAGTG	TCGCATGCAG	AACTTGTTAC	25680
	AGGTGTTAAT	GCTATTAAGC	AAAATGCTGA	TGCGTTAAAT	AATGcAATGG	GTACATTGAA	25740
	ACAACAAATT	CAAGCGAACA	GTCAAGTACC	ACAGTCAGTT	GACTTTACAC	AAGCGGATCA	25800
35	AGACAAACAA	CAAGCATATA	ACAATGCGGC	TAACCAAGCG	CAACAAATCG	CAAATGGCAT	25860
	ACCAACACCT	GTATTGACGC	CTGATACAGT	AACACAAGCA	GTGACAACCTA	TGAATCAAGC	25920
40	GAAAGATGCA	TTAAACGGTG	ATGAAAAATT	AGCACAAGCG	AAACAAGAAG	CTTTAGCAAA	25980
	TCTTGATACG	TTACGCGATT	TAAATCAACC	ACAACGTGAT	GCATTACGTA	ACCAAATCAA	26040
	TCAAGCACAA	GCGTTAGCTA	CAGTTGAACA	AACTAAACAA	AATGCACAAA	ATGTGAATAC	26100
45	aGCaATGAGT	AACTTGAAAC	aAGGTATTGC	aAACAAAGAT	ACTGTCAAAG	CAAGTGAGAA	26160
	CTATCATGAT	GCTGATGCCG	ATAAGCAAAC	AGCATATACA	AATGCAGTGT	CTCAAGCGGA	26220
	AGGTATTATC	AATCAAACGA	CAAATCCAAC	GCTTAACCCA	GATGAAATAA	CACGTGCATT	26280
50	AACTCAAGTG	ACTGATGCTA	AAAATGGCTT	AAACGGTGAA	GCTAAATTGG	CAACTGAAAA	26340
	GCAAAATGCT	AAAGATGCCG	TAAGTGGGAT	GACGCATTTA	AACGATGCTC	AAAAACAAGC	26400

55

	AGCAACGAGC CTAGATCAAG CAATGGATCA ATTATCACAA GCTATTAATG ATAAAGCTCA	26520
	AACATTAGCG GACGGTAATT ACTTAAATGC AGATCCTGAC AAACAAAATG CGTATAAACA	26580
5	GGCAGTAGCA AAAGCTGAAG CATTATTGAA TAAACAAAGT GGTACTAATG AAGTACAAGC	26640
	ACAAGTTGAA AGCATCACTA ATGAAGTGAA CGCAGCGAAA CAAGCATTAA ATGGTAATGA	26700
	CAATTTGGCA AATGCAAAAC AACAAAGCAAA ACAACAATTG GCGAACTTAA CACACTTAAA	26760
10	TGATGCACAA AAACAATCAT TTGAAAGTCA AATTACACAA GCGCCACTTG TTACAGATGT	26820
	CACTACGATT AATCAAAAAG CACAAACGTT AGATCATGCG ATGGAATTAT TAAGAAATAG	26880
	TGTTGCGGAT AATCAAACGA CATTAGCGTC TGAAGATTAT CATGATGCAA CTGCGCAAAG	26940
15	ACAAAATGAC TATAACCAAG CTGTAACAGC TGCTAATAAT ATAATTAATC AAACCTACATC	27000
	GCCTACGATG AATCCAGATG ATGTTAATGG TGCAACGACA CAAGTGAATA ATACGAAAGT	27060
20	TGCATTAGAT GGTGATGAAA ACCTTGACGC AGCTAAACAA CAAGCAAACA ACAGACTTGA	27120
	TCAATTAGAT CATTTGAATA ATGCGCAAAA GCAACAGTTA CAATCACAAA TTACGCAATC	27180
	ATCTGATATT GCTGCAGTTA ATGGTCACAA ACAACAGCA GAATCTTTAA ATACTGCGAT	27240
25	GGGTAACCTTA ATTAATGCGA TTGCAGATCA TCAAGCCGTT GAACAACGTG GTAACCTCAT	27300
	CAATGCTGAT ACTGATAAAC AAAGTCTTA TAATACAGCG GTAAATGAAG CAGCAGCAAT	27360
	GATTAAACAAA CAAAGTGGTC AAAATGCGAA CCAACAGAA GTAGAACAAG CTATTACTAA	27420
30	AGTTCAAACA ACACTTCAAG CGTTAAATGG AGACCATAAT TTACAAGTTG CTAAAACAAA	27480
	TGCGACGCAA GCAATTGATG CTTTAACAAG CTTAAATGAT CCTCAAAAAA CAGCATTAAA	27540
	AGACCAAGTT ACAGCTGCAA CTTTAGTAAC TGCAAGTTCAT CAAATTGAAC AAAATGCGAA	27600
35	TACGCTTAAC CAAGCAATGC ATGGTTAAG ACAGAGCATT CAAGATAACG CAGCAACTAA	27660
	AGCAATATAGC AAATATATCA ACGAAGATCA ACCAGAGCAA CAAAGCTATG ATCAAGCTGT	27720
40	TCAAGCCGCA AATAATATTA TCAATGAACA AACTGCAACA TTAGATAATA ATGCGATTAA	27780
	TCAAGCAGCG ACAACTGTGA ATACAACGAA AGCAGCATT CATGGTGATG TGAAGTTACA	27840
	AAATGATAAA GATCATGCTA AGCAAACGGT TAGTCAATTA GCACATCTAA ACAATGCACA	27900
45	AAAACATATG GAAGATACGT TAATTGATAG TGAACAACCT AGAACAGCAG TTAAGCAAGA	27960
	TTTGACTGAA GCACAAGCAT TAGATCAACT TATGGATGCA TTACAACAAA GTATTGCTGA	28020
	CAAAGATGCA ACACGTGCGA GCAGTGCTA TGTCATGCA GAACCGAATA AAAACAATC	28080
50	CTATGATGAA GCAGTTCAAA ATGCTGAGTC TATCATTGCA GGATTAAATA ATCCAATAT	28140
	CAATAAAGGT AATGTATCAA GTGCGACTCA AGCAGTAATA TCATCTAAAA ATGCATTAGA	28200

55

	TCAATTAACA CCAGCTCAAC AACAAAGCGCT AGAAAATCAA ATTAATAATG CAACAACCTCG	28320
	TGATAAAGTG GCTGAAATCA TTGCACAAGC GCAAGCATtA AATGAAGCGA TGAAAGCATT	28380
5	AAAAGAAAGT ATTAAGGATC AACCACAAAC TGAAGCAAGT AGTAAATTTA TTAACGAGGA	28440
	TCAAGCGCAA AAAGATGCTT ATACGCAAGC AGTACAACAC GCGAAAGATT TGATTAACAA	28500
	AACAACCTGAT CCTACATTAG CTAAATCAAT CATTGATCAA GCGACACAGG CAGTGACAGA	28560
10	TGCTAAAAAC AATTTACATG GTGATCAAAA ACTAGCTCAA GATAAGCAAC GTGCAACAGA	28620
	AACGTTAAT AACTTGTCTA ACTTGAATAC ACCACAACGT CAAGCACTTG AAAATCAAAT	28680
	TAATAATGCA GCAACTCGTG GCGAAGTAGC ACAAAAATTA ACTGAAGCAC AAGCACTTAA	28740
15	CCAAGCAATG GAAGCTTTAC GTAATAGCAT TCAAGATCAA CAGCAAACGG AAGCGGGTAG	28800
	CAAGTTTATC AATGAAGATA AaCCaCmAAA AGrTGCTTAC CAAGCAGCAG TTCAAAATGC	28860
20	AAAAGATTTA ATTAATCAAA CTAACAATCC AACGCTTGAT AAAGCACAAG TTGAACAATT	28920
	GACACAAGCT GTTAACCAAG CTAAGATAA CCTACACGGT GATCAAAAAC TTGCAGACGA	28980
	TAAACAACAT GCGGTTACTG ATTTAAATCA ATTAAATGGT TTGAATAATC CGCAACGTCA	29040
25	AGCACTTGAA AGCCAAATAA ACAACGCAGC AACTCGTGGC GAAGTAGCAC AAAAATTAGC	29100
	TGAAGCAAAA GCGCTTGATC AAGCAATGCA AGCATTACGT AATAGTATTC AAGATCAACA	29160
	ACAAACAGAA TCTGGTAGCA AGTTTATCAA TGAAGATAAA CCGCAAAAAG ATGCTTACCA	29220
30	AGCAGCAGTT CAAAATGCAA AAGATTTAAT TAACCAAACA GGTAATCCAA CACTCGACAA	29280
	ATCACAAGTA GAACAATTGA CACAAGCAGT AACAACTGCA AAAGATAATC TACATGGTGA	29340
	TCAAAAACCTT GCTCGTGATC AACACAAGC AGTAACAACCT GTAAATGCAT TGCCAACTT	29400
35	AAATCATGCA CAACAACAAG CATTAACTGA TGCTATAAAT GCAGCGCCTA CAAGAACAGA	29460
	GGTTSCACAA CATGTTCAAA CTGCTACTGA ACTTGATCAC GCGATGGAAA CATTGAAAAA	29520
40	TAAAGTTGAT CAAGTGAATA CAGATAAGGC TCAACCAAAT TACTACTGAAG CGTCAACTGA	29580
	TAAAAAAGAA GCAGTAGATC AAGCGTTACA AGCTGCAGAA AGCATTACAG ATCCAACTAA	29640
	TGGTTCAAAT GCGAATAAAG ACGCTGTAGA CCAAGTATTA ACTAAGCTTC AAGAAAAAGA	29700
45	AAATGAGTTA AATGGTAATG AGAGAGTCGC TGAAGCTAAA ACACAAGCGA AACAACTAT	29760
	TGACCAATTA ACACATTTAA ATGCTGATCA AATTGCAACT GCTAAACAAA ACATTGATCA	29820
	AGCGACGAAA CTTCAACCAA TTGCTGAATT AGTAGATCAA GCAACGCAAT TGAATCAATC	29880
50	TATGGATCAA TTACAACAAG CAGTTAATGA ACATGCTAAC GTTGAGCAAA CTGTAGATTA	29940
	CACACAAGCA GATTCAGATA AACAAAATGC TTATAAACAA GCTATTGCTG ATGCTGAAAA	30000

55

TGCAAAACAA GCATTAAATG GTGATGAACG TGTAGCACTT GCTAAAACAA ATGGTAAACA 30120  
 TGACATCGAC CAATTGAATG CATTAAACAA TGCTCAACAA GATGGATTTA AAGGTCGCAT 30180  
 5 CGATCAATCA AACGATTTAA ATCAAATCCA ACAAATTGTA GATGAGGCTA AGGCACTTAA 30240  
 TCGTGCAATG GATCAATTGT CACAAGAAAT CACTGACAAT GAAGGACGCA CGAAAGGTAG 30300  
 CACGAACTAT GTCAATGCAG ATACACAAGT CAAACAAGTA TATGATGAAA CGGTTGATAA 30360  
 10 AGCGAAACAA GCACTTGATA AATCGACTGG TCAAACTTA ACTGCAAAAC AAGTTATCAA 30420  
 ATTAAATGAT GCAGTCACTG CAGCTAAGAA AGCATTAAAT GGTGAAGAAA GACTTAATAA 30480  
 TCGTAAAGCT GAAGCATTAC AAAGATTGGA TCAATTAACA CATCTAAACA ATGCTCAAAG 30540  
 15 ACAATTAGCA ATCCAACAAA TTAATAATGC TGAAACGCTA AATAAAGCAT CTCGAGCAAT 30600  
 TAATAGAGCA ACTAAATTAG ATAATGCAAT GGGTTCAGTA CAACAATATA TTGACGAACA 30660  
 GCACCTTGGT GTTATCAGCA GCACAAATTA CATCAATGCA GATGACAATT TGAAAGCAAA 30720  
 20 TTATGATAAT GCAATTGCGA ATGCAGCACA TGAGTTAGAT AAAGTGCAAG GTAATGCAAT 30780  
 TGCaaaAGCT GAAGCAGAGC AATTGAAACA AAATATTATC GATGCTCAA ATGCATTAAA 30840  
 25 TGGAGACCAA AACCTTGCAA ATGCCAAAGA TAAAGCAAAT GCGTTTGTTA ATTCGTTAAA 30900  
 TGGATTAAAT CAACAGCAAC AAGATCTTGC ACATAAAGCA ATTAACAATG CCGATACTGT 30960  
 ATCAGATGTA ACAGATATTG TTAATAATCA AATTGACTTA AATGATGCAA TGGAAACATT 31020  
 30 GAAACATTTA GTTGACAATG AAATCCAAA TGCAGAGCAA ACTGTCAATT ACCAAAACGC 31080  
 TGACGATAAT GCTAAA 31096

## (2) INFORMATION FOR SEQ ID NO: 60:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2243 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

45 ATGACAGAAT GGGAGCGAGG ACTTAGAATG TTTCCTAAAT CAGGTTTATT AAATTTTGAG 60  
 TTAGCGATAG mAAATCGTTC ATTAAATGAT GATGAAAAAG CATTAAAATA TGTGCGTAAA 120  
 GCATTAAATG CAGACCCTAA AAATACAGAT TATATTAAC TAAAAAGA GTTGAATAA 180  
 50 TCAATGAGT CGAAAAATAA ATAACCTTTA TGATGTACAA CAGTTATTGA AAAGTTACGG 240  
 ATTTCTAATA TATTTTAAAA ATCCAGAAGA TATGTACGAA ATGATTCAAC AGGAGATTTC 300

55

	TAATCAGAGA AGGAATGAAC AGAAATGACA AAAATTATTT TAGCAGCTGA TGTAGGCGGG	420
	ACGACTTGTA AATTAGGTAT TTTCACACCT GAATTAGAAC AATTACATAA ATGGTCTATT	480
5	CACACTGATA CATCTGATAG TACAGGATAT ACACTTTTGA AAGGAATTTA TGATTCGTTT	540
	GTTGAAAAAG TAAATGAAAA TAATTATAAT TTTTCAAATG TACTTGGCGT AGGTATTGGT	600
	GTACCAGGTC CTGTTGACTT TGAAAAAGGT ACAGTAAATG GAGCAGTAAA CTTATATTGG	660
10	CCAGAAAAAG TTAATGTACG TGAGATTTTT GAACAATTCG TTGATTGTCC AGTGTATGTA	720
	GATAATGATG CTAACATAGC TGCTTTAGGG GaGAAACACA AAGGTGCTGG TGAAGGTGCC	780
15	GATGATGTTG TTGCCATCAC ACTTGGTACA GGTCTAGGTG GAGGAATTAT TTCCAAATGG	840
	TGAAATCGTA CATGGTCATA ATGGCTCtGG CGCAGAAATA GGTCATTTTA GAgCAGACTT	900
	CgATCAACGA TTtaATGTA ATTGTGGTCG TTCTGGATGT ATTGAAACAG TTGCTTCaGC	960
20	GACAGGCGTT GTTAACTTAG TTAACCTtCa CTATCCGAAG TTGACGTTTA GATCTTCTAT	1020
	ATTAGAATTG ATTAAAGAAA ATAAGGtTAC aGCAAAAGCT GTTTTTGATG CGGCAAAAGC	1080
	TGGTGACCAA TTCTGTATTT TCATTACTGA AAAGGTGCA AACTATATTG GATATTTATG	1140
25	TAGTATTATT AGTGTACAA GTAATCCGAA ATATATCGTT CTAGGTGGAG GAATGTCTAC	1200
	TGCAGGACCT ATTTTAATTG AAAATATTAA AACAGAATAT CATAATTTAA CATTTGCACC	1260
	TGCTCAATTT GAAACTGAAA TTGTACAAGC GAAATTAGGT AATGATGCAG GTATTACAGG	1320
30	AGCAGCAGGA TTAATCAAGA CCTATGTATT AGATAAAGAG GGGGTAAAAT AATGGCTATT	1380
	GTTGATGTGG TTGTTATTCC AGTTGGAACG GAAGGTCCGA GTGTTAGTAA ATATATTGCA	1440
	GATATTCAGA AAAAaCTTCA AGAATATAAA GCAATGGGTA AAATTGATTT TCAATTAACA	1500
35	CCAATGAATA CTCTAATTGA AGGTGAATTA AGCGATGTAT TAGAAGTTGT GCAAGTGATA	1560
	CATGAATTAC CTTTTGATAA AGGTTTAAGT AGAGTTTGTA CAAATATCCG TATTGATGAC	1620
	CGACGAGACA AATCTAGAAA AATGAATGAT AAaCTAACAT CAGTACAAAA ACATTTAGAA	1680
40	AATAGTGGTG AAAACCTATG AGGATTTCAA GCTTAACTTT AGGCTTAGTT GATACTAATA	1740
	CGTATTTcAT CGAAAATGAC AAAGCTGTTA TTCTGATTGA CCCTTCAGGT GAAAGTGAAA	1800
45	AAATTATTAA AAAATTAAAC CAAATAAATA AACCgTTAAA AGCTATTTTA TTAACACATG	1860
	CACACTTTGA TCATATCGGA GCAGTCGATG ATATAGTTGA TCGATTcGAT GTCCCGGTTT	1920
	ATATGCATGA AGCAGAGTTT GATTTTCTAA AAGATCCCGT TAAAAATGGG GCAGATAAAT	1980
50	TTAAGCAATA TGGATTACCA ATTATTACAA GTAAGGTAAC TCCTGAAAAG TTAAmCGAAG	2040
	GTAGCACAGA AATAGAAGGA TTTAAGTTnT nAyrTGtCa CACACCTGGA CATTcACCAG	2100

55

GAATCGGACG TACAGATTTA TATAAAGGTG ATTATGAAAC GCTAGTTGAT TCTATTCAAG 2220  
ATAAAATATT TGAATTAGAA GGC 2243

5 (2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8009 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

15 TTGGnATCAT tyAcgGTAAa AAGAATAaaG CAAGATTtAT TTCATTAGTA CTAATTTGTG 60  
 CAATGTTTGC AATTTGTTGG GTTGCATATA TTCAATGGGA GTCTACAATC GCTTCATTTA 120  
 20 CACAATCTAT TAATATTTTca ATGGCACAAAT ATAGTGTTTT ATGGACAATT AACCGAATAA 180  
 TGATTTTAGT AGCACAAACCA TTAATTAAAC CGATTCTCTA TCTGTTAAAA GGAAACTTAA 240  
 AGAAGCAAAT GTTTGTCGGC ATCATCATTT TTATGTTGTC GTTCTTTGTC ACGAGTTTGT 300  
 25 CCGAAAACCTT TACAATATTT GTTGTCTGGTA TGATTATTTT AACTTTGTGA GAAATGTTTG 360  
 TATGGCCAGC AGTTCCAAC TATAGCCAATC AGTTAGCGCC AGATGGTAAG CAAGGACAGT 420  
 ACCAAGGTTT TGTGAATTCA GCTGCTACAG TAGGAAAAGC ATTTGGTCCA TTTCTTGGTG 480  
 30 GTGTATTAGT TGATGCGTTT AATATGCGCA TGATGTTTAT CGGTATGATG CTACTACTTG 540  
 TATTTGCATT AATATTATTA ATGGTTTTCA AGGAGAATAA TACGCAACCT AAAAAAATAG 600  
 ATGCATAATG AGTAAATAGA ATTAACGTTA TAGACTTGAA ATAAATGTCG TTATAACATA 660  
 35 ATATTAATTT GTATAATTTA ATTTCTTTTG GAGCTTTTCT ACAGAAAGCT AGTGATGCTG 720  
 AGAGCTAGTG TTAAGGACTA AATGTAAATC GTATTAAATTT TAAATTGAAT GAATGACATC 780  
 TCTTACTATT AAAATGAGTG CACAATTTTT GTGAAATAGG GTGGTAACGC GGCAAATGTC 840  
 40 GTCCCTATGT AAATAGAATA GTTAGAGGTG TCTTTTTTAT TGAATAGGAG GAAATGTGTT 900  
 GAATTACAAC CACAATCAAA TTGAAAAGAA ATGGcAAGAC TATTGGGACG AAAATAAAAC 960  
 45 ATTTAAAACA AATGATAACT TAGGTCAAAA GAAATTTTAT GCTTTAGACA TGTTTCCATA 1020  
 TCCATCAGGT GCTGGTTTAC ATGTTGGACA TCCTGAGGGc TATACAGCAA CAGATATCAT 1080  
 TTCAAGATAT AAAAGAATGC AAGGATATAA TGTATTACAT CCGATGGGGT GGGATGCATT 1140  
 50 CGGATTACCA GCAGAGCAAT ATGCTTTAGA CACTGGCAAC GACCCACGTG AATTTACAAA 1200  
 GAAAAATATC CAAACTTTTA AACGACAAAT TAAAGAATTA GGGTTCAGTT ATGATTGGGA 1260

55

EP 0 786 519 A2

	GTATATAAC AAAGGTTTAG CATACGTTGA TGAAGTTGCA GTTAACTGGT GTCCAGCATT	1380
	AGGCACTGTT TTATCTAACG AAGAAGTGAT TGATGGTGTC TCTGAACGTG GTGGACATCC	1440
5	AGTTTATCGT AAGCCGATGA AACAAATGGGT ACTTAAAATC ACAGAATATG CAGATCAATT	1500
	ATTAGCAGAT TTAGATGATT TAGATTGGCC TGAGTCTTTA AAAGATATGC AGCGCAATTG	1560
	GATTGGACGT TCTGAAGGGG CCAAAGTTTC ATTTGATGTA GATAATACGG AAGGAAAAGT	1620
10	AGAAGTATTT ACGACTAGAC CAGATACAAT CTATGGTGCA TCATTCTTAG TCTTAAGTCC	1680
	TGAACATGCA TTAGTTAATT CAATTACAAC AGATGAATAT AAAGAAAAAG TAAAAGCTTA	1740
15	TCAACAGAA GCTTCTAAAA AGTCAGATTT AGAACGTACA GATTTAGCAA AAGATAAATC	1800
	AGGTGTATTT ACTGGTGCAT ATGCAACTAA TCCTTTATCT GGTGAAAAAG TACAAATTTG	1860
	GATTGCTGAT TATGTATTAT CAACATATGG TACTGGAGCA ATTATGGCAG TACCAGCGCA	1920
20	TGATGACAGA GATTATGAAT TTGCTAAAAA GTTTGATTTG CCAATCATTG AAGTCATCGA	1980
	AGGTGGAAT GTTGAAGAAG CAGCATACAC TGGTGAAGGT AAACATATTA ATTCTGGTGA	2040
	ACTTGATGGT TTAGAAAATG AAGCGGCAAT TACTAAAGCT ATTCAATTAT TAGAGCAAAA	2100
25	AGGTGCTGGC GAAAAGAAAG TTAATTACAA ATTAAGAGAT TGGTTATTCA GTCGTCAGCG	2160
	TTATTGGGGC GAACCAATTC CTGTCATTCA TTGGGAAGAT GGAACAATGA CAACTGTTCC	2220
	TGAAGAAGAG CTACCATTGT TGTTACCTGA AACAGATGAA ATCAAGCCAT CAGGGACTGG	2280
30	TGAGTCTCCA CTAGCTAATA TTGATTCATT TGTAATGTT GTAGATGAAA AAACAGGTAT	2340
	GAAAGGACGT CGTGAAACAA ATACAATGCC ACAATGGGCA GGTAGTTGTT GGTATTATTT	2400
	ACGTTACATC GATCCTAAAA ATGAAAATAT GTTAGCAGAT CCTGAAAAAT TAAAACATTG	2460
35	GTTACCTGTT GATTTATATA TCGGTGGAGT AGAACATGCG GTTCTTCACT TATTATATGC	2520
	AAGATTTTGG CATAAAGTCC TTTATGATTT GGCTATCGTA CCTACTAAAG AACCTTTCCA	2580
40	AAAATTATTT AACCAAGGTA TGATTTTAGG AGAAGGTAAT GAGAAGATGA GTAAATCTAA	2640
	AGGAAATGTA ATCAATCCTG ATGATATAGT ACAGTCTCAT GGTGCAGATA CTTTGCGTCT	2700
	TTACGAAATG TTTATGGGAC CTTTAGATGC TGCAATTGCA TGGAGTGAAA AAGGATTAGA	2760
45	TGGGTCTCGT CGATTCTTAG ATCGCGTATG GCGTTTAATG GTAAATGAAG ATGGGACATT	2820
	GAGTTCAAAA ATTGTAACTA CAAATAATAA ATCTTTAGAT AAAGTTTATA ACCAACTGT	2880
	TAAAAAGGTA ACAGAAGACT TTGAAACATT AGGATTTAAT ACTGCTATTA GTCAATTAAT	2940
50	GGTATTTATT AATGAGTGTT ATAAAGTTGA TGAAGTTTAT AAACCTTACA TTGAAGGCTT	3000
	CGTTAAAATG TTAGCACCTA TTGCACCACA TATCGGTGAA GAATTATGGT CAAAATTAGG	3060

55

	TGATGAAGTA	GAAATCGTTG	TTCAAGTGAA	TGGTAAATTG	AGAGCTAAAA	TTAAAATTGC	3180
	TAAAGATACA	TCAAAAGAAG	AAATGCAAGA	AATTGCCTTA	TCTAATGACA	ATGTTAAAGC	3240
5	GAGTATTGAA	GGTAAAGACA	TCATGAAAGT	CATCGCTGTT	CCTCAAAAAT	TAGTCAATAT	3300
	TGTAGCTAAA	TAATGTTTTA	AGGAGGACTT	TGAAATGAAG	TCAATTACTA	CAGATGAATT	3360
	AAAAAATAAA	CTTTTAGAAT	CTAAACCAGT	TCAAATTGTT	GATGTTCGTA	CTGATGAAGA	3420
10	AACAGCAATG	GGATATATTG	CTAATGCAAA	GTTAATTCCA	ATGGATACCA	TTCCGGATAA	3480
	TTTAAATTCA	TTTAATAAAA	ATGAAATATA	TTATATTGTA	TGTGCTGGTG	GAGTTCGAAG	3540
	CGCTAAAGTT	GTAGAATATT	TAGAGGCAAA	TGGCATTGAT	GCCGTAAATG	TCGAAGGCGG	3600
15	CATGCACGCA	TGGGGCGATG	AAGGTTTGGG	AATAAAAAGT	ATTTAAAGTA	GTGACATAAT	3660
	TTAAAATAAT	ATTACATTTG	TAATGACACC	AAGTAACGTT	TCGGTTGCTT	GGTGTTTTTT	3720
20	GGTATGAATT	ACTTCTGTT	ACAAAACAAT	CTAAAGCGTT	CTTGTTATGT	TTTATTAAAG	3780
	TTTTAATTAC	AAAACGGAAA	CTAAATTGTA	ATAAAATAAA	ACTTTATTTT	ATAAAATGAT	3840
	GATGATAAAA	TTGAGTGAAC	TTAAAATATT	GTACAAAATA	ATATAGCTAT	AAATATAATA	3900
25	TAGCTATAAA	TATAATATGA	GGGAGCGTAT	ATTTTGTAGCA	TAATTCTTAA	CAACACAGCA	3960
	GAGAACAGAC	AACCAGGAGG	AAAATGAAAT	GAATTTGTTA	AAGAAAAATA	AATATAGTAT	4020
	TAGGAAGTAT	AAAGTAGGCA	TATTCTCTAC	TTTAATCGGA	ACAGTTTTAT	TACTTTCAAA	4080
30	CCCAAATGGT	GCACAAGCCT	TAACTACGGA	TAATAATGTA	CAAAGCGATA	CTAATCAAGC	4140
	AACACCTGTA	AATTCACAAG	ATAAAGATGT	TGCTAATAAT	AGAGGTTTAG	CAAATAGTGC	4200
	GCAGAATACA	CCTAATCAAT	CTGCAACAAC	CAATCAAGCA	ACGAATCAAG	CATTGGTTAA	4260
35	TCATAATAAT	GGTAGTATAG	TAAATCAAGC	TACGCCAACA	TCAGTGCAAT	CAAGTACGCC	4320
	TTCAGCACAA	AACAATAATC	ATACAGATGG	CAATACAACA	GCAACTGAGA	CAGTGTCAAA	4380
	CGCTAATAAT	AATGATGTAG	TGTCGAATAA	TACCGCATTG	AATGTACCAA	CTAAAACAAA	4440
40	TGAAAATGGT	TCAGGAGGAC	ATCTAACTTT	AAAGGAAATT	CAAGAAGATG	TTCGTCATTC	4500
	TTCAAATAAA	CCAGAGCTAG	TTGCAATTGC	TGAACCAGCA	TCTAATAGAC	CGAAAAAGAG	4560
45	AAGTAGACGT	GCGGCACCGG	CAGATCCTAA	TGCAACTCCA	GCAGATCCAG	CGGCTGCAGC	4620
	GGTAGGAAAC	GGTGGTGCAC	CAGTTGCAAT	TACAGCGCCA	TATACGCCAA	CAACTGATCC	4680
	TAATGCCAAT	AATGCAGGAC	AAAATGCACC	TAACGAAGTG	CTGTCATTTG	ATGACAATGG	4740
50	TATTAGACCA	AGTACCAACC	GTTCTGTGCC	AACAGTAAAC	GTTGTTAATA	ACTTGCCGGG	4800
	CTTCACACTA	ATCAATGGTG	GCAAAGTAGG	GGTGTTTAGT	CATGCAATGG	TAAGAACGAG	4860

55



TCGTATACAT GGAAGTATA CGAATGACCA TGGCGATTTT AATGGTATCG AGAAAGCATT 4980  
 AACAGTAAAT CCGAATTCTG AATTAATCTT TGAATTTAAT ACAATGACTA CTAAAAACGG 5040  
 5 TCAAGGCGCA ACAAATGTTA TTATCAAAAA TGCTGATACT AATGATACGA TTGCTGAAAA 5100  
 GACTGTTGAA GCGGTCCTCA CTTTGCCTTT ATTTAAAGTA CCTGATAATG TGAGAAATCT 5160  
 CAAAATTCAA TTTGTACCTA AAAATGACGC AATAACAGAT GCGCGTGGCA TTTATCAACT 5220  
 10 AAAAGATGGT TACAAATACT ATAGCTTTGT TGAATCTATC GGACTTCATT CTGGGTCACA 5280  
 TGTTTTTGTT GAAAGACGAA CAATGGATCC AACAGCAACA AATAATAAAG AGTTTACTGT 5340  
 AACACATCA TTAAAGAATA ATGGTAATTC TGGTGCTTCT CTAGATACAA ATGACTTTGT 5400  
 15 ATATCAAGTT CAATTACCTG AAGGTGTTGA ATATGTGAAC AATTCATTGA CTAAAGATTT 5460  
 TCCAAGTAAC AATTCAGGCG TTGATGTTAA TGATATGAAT GTTACATATG ATGCAGCAAA 5520  
 TCGTGTGATA ACAATTAAAA GTACTGGAGG AGGTACAGCA AACTCTCCGG CACGACTTAT 5580  
 20 GCCTGATAAA ATACTCGATT TAAGATATAA ATTACGTGTA AATAATGTGC CGACACCAAG 5640  
 AACAGTAACA TTAAACGAGA CATTAACTGA TAAACATAT ACACAAGATT TCATTAAATC 5700  
 25 AGCTGCAGAA AGTCATACTG TAAGTACAAA TCCATATACT ATCGATATCA TCATGAATAA 5760  
 AGATGCATTA CAAGCCGAAG TTGACAGACG TATTCAACAA GCTGATTATA CATTTGCGTC 5820  
 ATTAGATATC TTAAATGGTC TGAAACGACG CGCACAAACG ATTTTAGATG AAAATCGTAA 5880  
 30 CAATGTACCA TTAAATAAAA GAGTTTCTCA AGCATATATT GATTCATTAA CTAATCAAAT 5940  
 GCAACATACG TTAATTCGAA GTGTTGATGC TGAAAATGCA GTTAATAAAA AAGTTGACCA 6000  
 AATGGAAGAT TTAGTTAATC AAAATGATGA ATTGACAGAT GAAGAAAAAC AAGCAGCAAT 6060  
 35 ACAAGTTATC GAGGAACATA AAAATGAAAT AATTGGTAAT ATTGGTGACC AAACGACTGA 6120  
 TGATGGCGTT ACTAGAATCA AAGATCAAGG TATACAGACC TTAAGTGGGG ATACTGCAAC 6180  
 ACCGTTTGTT AAACCAAATG CTAAAAAAGC AATACGTGAT AAAGCAACGA AACAAAGGGA 6240  
 40 AATTATCAAT GCAACACCAG ATGCTACTGA AGACGAGATT CAAGATGCAC TAAATCAATT 6300  
 AGCTACGGAT GAAACAGATG CTATTGATAA TGTACGAAT GCTACTACAA ATGCTGACGT 6360  
 TGAAACAGCT AAAAATAATG GCATCAATAC TATTGGAGCA GTTGTTCCTC AAGTAACTCA 6420  
 45 TAAAAAGCT GCAAGAGATG CAATTAACCA AGCAACAGCA ACGAAAAGAC AACAAATAAA 6480  
 TAGTAATAGA GAAGCAACTC AGGAAGAGAA AAATGCAGCA TTGAACGAAT TAACTCAAGC 6540  
 50 AACCAACCAT GCTTTAGAAC AAATCAATCA AGCAACAACA AATGCTAATG TTGATAACGC 6600  
 CAAAGGAGAT GGTCTAAATG CCATTAATCC AATTGCTCCT GTAAGTGTG TTAAGCAAGC 6660

55

EP 0 786 519 A2

TGATGCGACT CAAGAAGAAA GACAAGCAGC AATTGACAAA GTGAATGCTG CTGTAAGTGC 6780  
 AGCAAACACA AACATTTTAA ACGCTAATAC CAATGCTGAT GTTGAACAAG TAAAGACAAA 6840  
 5 TGCGATTCAA GGAATACAAG CAATTACACC AGCTACAAAA GTAAAAACAG ATGCAAAAAA 6900  
 TGCCATCGAT AAAAGTGCGG AACGCAACA TAATACGATA TTTAATAATA ATGATGCGAC 6960  
 GCTCGAAGAA CAACAAGCAG CACAACAATT ACTTGATCAA GCTGTAGCCA CAGCGAAGCA 7020  
 10 AAATATTAAT GCAGCAGATA CGAATCAAGA AGTTGCACAA GCAAAAGATC AGGGCACACA 7080  
 AAATATAGTA GTGATTCAAC CGGCAACACA AGTTAAAACG GATACTCGCA ATGTTGTAAG 7140  
 TGATAAAGCG CGAGAGGCGA TAACAAATAT CAATGCTACA ACTGGCGCGA CTCGAGAAGA 7200  
 15 GAAACAAGAA GCGATAAATC GTGTCAATAC ACTTAAAAAT AGAGCATTAA CTGATATTGG 7260  
 TGTGACGTCT ACTACTGCGA TGGTCAATAG TATTAGAGAC GATGCAGTCA ATCAAATCGG 7320  
 CGCAGTTCAA CCGCATGTAA CGAAGAAACA AACTGCTACA GGTGTATTAA ATGATTTAGC 7380  
 20 AACTGCTAAA AAGCAAGAAA TTAATCAAAA CACAAATGCA ACAACTGAAG AAAAGCAAGT 7440  
 GGCTTTAAAT CAAGTGGATC AAGAGTTAGC AACGGCAATT AATMATATAA ATCAAGCTGA 7500  
 25 TACAAATGCG GAAGTAGATC AAGCGCAACA ATTAGGTACA AAAGCAATTA ATGCGATTCA 7560  
 GCCAAATATT GTTAAAAAAC CTGCAGCATT AGCACAAATC AATCAGCATT ATAATGCTAA 7620  
 ATTAGCTGAA ATCAATGCTA CACCAGATGC AACGAATGAT GAGAAAAATG CTGCGATCAA 7680  
 30 TACTTTAAAT CAAGACAGAC AACAAAGCTAT TGAAAGTATT AAACAAGCTA ACACAAATGC 7740  
 AGAAGTAGAC CAAGCTGCGA CAGTAGCAGA GAATAATATC GATGCTGTTC AAGTTGATGT 7800  
 AGTAAAAAAA CAAGCAGCGC GAGATAAAAT CACTGCTGAA GTGGcGAacG TATTGaAGCG 7860  
 35 GTTAAACAAA CACCTAATGC AACTGACGAA GAAAAGCAGG CTGCTGTTAA TCAAATCCAA 7920  
 TCAACTTTAA AGATTCAAGC AATTTAATCC AATTTAATC CAAAACCCAA ACAAATGGAT 7980  
 TCAGGGTAGG ACACCACTTA CAAATCCAA 8009

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10953 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

ACCCACCCcN TGGGGATAnT TTACCTGGTG GGGCCTTCGA TTGCCTTTAG GTGAAACCaG 60

EP 0 786 519 A2

	AGATGAATGC TAACCATATT CATTCTGCTA AAGATGGTCG TGTTACTGCG ACAGCTGAAA	180
	TTATTCATCG AGGTAAGTCG ACACATGTAT GGGATATAAA AATTAAGAAT GACAAAGAAC	240
5	AATTAATTAC AGTTATGCGT GGTACAGTTG CTATTAAACC TTTAAAATAA AAGAACTGCT	300
	AGCTGAAATG TTATGAGATA TTCATAACTA CGGCTAGCAG TTTTTTTATG CGCTATATTG	360
	TTGTAGTTTT AGAAATGCTT GTTCAATGCG TTCGGCAGCT TTACGGCCAC CCATAACATT	420
10	TCTACCAAAT GGTCCTAATT CTAAGTCTGC AAAGCATCCT GCGACAAATA GATTTGGTAT	480
	CCATTCTAAT TTTTCGGAAA TAACAGGGTA ATTACATTCG TTGATAGGTG CATCATAATT	540
15	TTGTATTAAT TGCTTAATAA GTGGTTGTGA CATAAAATCT TGTTCAAAAC CAGTTGCAAC	600
	CATAATCTGT TGATATGGAA CAGAATCATT TTCAGTGTTA ATTACACCAC CACTAATTTG	660
	AGTGATAGGT GTTTTATGCa CATTATACG ACCATTTTTA ATATGTTTTT TAAGGCGTAA	720
20	GTACAGTTCG TGAGGCATTG ATCCTTTATG ACGTTCGCGT TGTACAATGG CATTTCTTTC	780
	AGGCATGCTT TTAGTACTTA AAAATGAAGA CATATTTTTT GACCTAACC AACCAGGATC	840
	AGCATCAAAG TCATGTATTT CAATATCTTT ATTTAGCCAT AAATGAATCT TTTTATCGTT	900
25	ATCATGATTT AACAAITTA A GTGCAAGATG TGCAGCAGTa ATGCCGCTAC CAACGATATG	960
	ATCGGTCTTA TCATATACTA CTTGATCAAG TTCTTTCTCG AAGATATGAT TTACATTCTG	1020
	TTTGTCTTTT AAAATGTCAG GCATAAACGG AATATTTGTA CTGCCTATTG CAATAACGAC	1080
30	GCAATCTGTA GTGATAATTT GTCCATCTTC TAACTTGATA TGCCATTGTG CTCTCTGTTT	1140
	ATCTAAAGTT TGAACATAAC CTTGAACCAA GCAATCCTCT AATTGATATT GTTTAGAAGC	1200
	ATGTGCAATA TGATCCATAA ACATTGTCAA TTCAGGTCGT TGATAAGGAC CATAAAAAGC	1260
35	ATTTGTATAT TGGTGCTGTT TAGCGAATTG TTTTAGATGG AACGGTTGTG GATGTACGTG	1320
	ATGTACAATC GGTGATCTTA AATAAGGCAT TTCTATTCTGA TTTGTATATG AGTTAAACCT	1380
40	TTGGCAAAAA GTTTCGTGTG GGTCAATGAT TGTTAATCGG TCTGTTGTGA ATCCGCTTGA	1440
	TAATAGTTTT TGTGCGATTG CAGTTCCTTG TATGCCACCG CCGATAATTG TCCAATGCAT	1500
	AATAAACCT CTCTCTTTTT AAAACGTAAT AGTTACGATT TATAATTATT ATTATCATAA	1560
45	TACATAACGA CATGAAAGGC AATTAAATTA AAGAGATATA TGTAGATAGG GCGAATCTGT	1620
	AGTCAAAGAA AAAATCATTG AAAAAGAGGT AACAAATGCA AAAGAWAACA GCAGTAAAT	1680
	CATTCTAAT TTGGAATCAT CTTACTGCTG TTTGTTGTG ATTTATATTC ATGATTTTGT	1740
50	TATATAATCT ACAATTTTGT GTCTTTTAAG TCTTCCGAAA TTTCATCGAC TTTAGTCTTT	1800
	TTAGTATAAG GCGTTTTAAT ATTATATGCT GCTTTCATAA TCATATGACT TGAAAGAGGA	1860

55

	GCAATAAAAT ATAAAAACGT ACCAAATAGT AATGACATTG CACCTAATGT TGATGCTTTT	1980
	CCGGCAGCAT GTGCACGTGA ATATACATCT TCAAGTCTCA ATAATCCTAT AGCTGCTAGG	2040
5	GCGCTAATTA AAGCACCGAT GATAACAAAG ATAAGTGCAA GACTAATCAG TATGATTTTG	2100
	ATCATGTTCA ATCACCTTAC CTTTGTCCAT AAATTIAGAG AATACTGCAG TACCTAAAAA	2160
	AGCTAATATA CCAATCATCA TAATAACGAC AATCATGTAT TTAATATTTA ATAAAATACT	2220
10	GAATAATGCT ATAAGTCCCA TTAATTGAAG ACCAATCGCA TCTAATGCGA CAACACGATC	2280
	GGCAAGTGAT GGGCCTAGCA CAACGCGAAT GAGCATAGCT AACATAGAAA TGACAACATAT	2340
15	GATTAATGCA ATAACGATAA TAACATTATG ATTCAATTATA TTTGCCCCAC CTCTCTTACA	2400
	ATTTTCTCTA ATGATGTTTT AATACTTTCT ACTTCTTGCT CTTTAGTTGA AAAATCTATG	2460
	GCATGAATAT AAATTTTTGT ACGATCGTCA CTTACACCAA GCACTACAGT ACCAGGTGTT	2520
20	AATGTAATTA AATTAGACAG CAAGACAATT TGCCAATCTT TTTTAAATC TGTGTGATAA	2580
	ACAAAGAATC CTGGTTCATT TTTAATCGAA GGTTAATAAA TAATTTTCAA AACATCAAAA	2640
	TTAGCTTTAA TCAGTTCGAT TAAGAAAATA ATAACATAAT TAATAATACG ATATAGCGTG	2700
25	ATGACATAAA ATCTACCTGG TAACACTCTG TGTAAGAGGT AAACAAGAAC TAGGCCAAAG	2760
	ATGAAACCTA ACACAAAGTT ATTTGTTGTG TAACTATTTG TCACAAACAA CCAAAACACT	2820
	GCGATAATAA AGTTTAATAC TAATTGTACA GCCATGTTAT TTACCTCCTA ATACAGCTTT	2880
30	AACGTAGGTT GATGGATTGT AGAATGTTTC TGCACCAGCT TTTACCATTG GATATAAGTA	2940
	ATCTGCTGAC AATCCATATA AAACAGTTAT CACAACCTGCA ACGATTGCAA TCGTAGTTAA	3000
	ATATTTGACG TCGACTTTGT TATTAAGATC ATATCCTTTT GGTGACCGA AAAAGCCTTG	3060
35	TAGGAATATG CGAATGACAG AATATAATAC GACTAAACTT GATAATAAGA CGATGACACC	3120
	ACTTAAATAA AATCCTCTTT CAAATGTTGA TTGGACAATA AAAAATTTTC CATAAAAGCC	3180
40	ACTGAGTGGG GGAATGCCAG CTAAACTTAA TGCTGCGATA AAGAATGACC AACAAGTAC	3240
	AGGATATCGT TTAATTAAGC CACCAAATTG TCTTAAATCA GCAGTGCCTG TAATTTTAAT	3300
	CATAATTCCG ATAAGCAAGA ATAATGCAAG TTTTACTAAC ATGTCGTGCA ATGTATAGTA	3360
45	AATAGCCCCA ATCATACCTG ACTCTGTCAT CATTGCAACG CCGACTAAGA TCACACCTAC	3420
	AGCAATCATG ACATTGTATA GGATGATTTT TTTAATGTTG GCATATGCAA CAGCACCGAC	3480
	ACAACCAAAG ATGATCGTTA ATAGTGCTAA GAATAAAATG ACATAATGTG AAAAGCTTAC	3540
50	ATTATCACTA AAGAATAGGC TCAATGTTCT AGCGATTGCA TAAACACCAA CTTTGTGTTAA	3600
	CAAAGCACCA AAGAATGCAA TGATTGGAAT TGGTGGgCAT AGTATGCACT AGGTAACCAA	3660

55

	ATATTGACTA AGCCACTGTC ATGCGCTGAA AGGTTAGCTA ATTTATTGCT TATATCTGCT	3780
	AGATTCAATG TTCCTACTAC TGAATATAAA ATCGCTACAC CCATTACGAA GAAGGATGAC	3840
5	GATACAACGT TAACAAGAAC ATATTTTATT GTTCTTGTA GTTGAATTTT TGTAGAACCA	3900
	ATTACTAATA AGAAATAAGA TGACATTAAA AATACTTCGA AAAATACGAA TAGGTTGAAA	3960
	ATGTCACCAG TTGTGAATGC ACCAATGATA CCTATTAACA TAAATAGTAC TGAAAAATAA	4020
10	TAATAATATC TTTCACGTTC AATACCAATT GTTTGGTATG AATATAAAAT CACAATAGCT	4080
	GTAATAATAA TACTAGTAAT TATTAGTAGG GCACTGAATA TGTCTAATAC AAAGACAATA	4140
	CTGTATGGTG CTTTCCATGA ACCTAGCTCT ACGCGTATTG GTCCATGTTT AACACATTT	4200
15	GCTAAATTGA TAATTGCCGC GACCAAGGTT AATAATGTAC CGCCTAGTGC GACATAACGC	4260
	TTTATAATAG GACGCTTTCC AATAAAGACA AGTAATATGG CTGTAATTAC TGGAATAACT	4320
	AGCGTTAACA CAAGCATATT ACTTTCAATC ATCTTCTGGA ACTCCTTTCA TACTCTCAAC	4380
20	GTTATCTGTG CCTAATTCTT TATATGTTCT AAATGCTAAT ACTAAGAAAA AGGCTGTTGT	4440
	CGCAAgGCGA TAACGATTGC TGTAAAATA AGTGCTTGCG GGaTAGGaTC AACATAGCTT	4500
25	TTTACGTTTC CTTCATAAAT TGGAACAGTA CCATGTTTAA GTCCGCCCAT AGTTATTAAA	4560
	AATAAATTG CTGCATGTGT TAATAGTGTA GTTCCCATAA CAATTCGTAT CAGACTTTTA	4620
	GACAAAACGA GATAGACACT AATTGCTGTG AGAATACCAC TAACAAAAAT CATAATAATT	4680
30	TCCACTATTC GTTCTCTCCA ATCGAAATAA TAATTGTCAT GACAGTACCA ACTACTGCAC	4740
	ATAAACACC GAAATCAAAG AATACTGCTG TTGTCATATG AACAGGTTCT AATATAAATA	4800
	ACGGTATATC AAATGTGACA TCGGTAAAGA AATTTTTGCC TAAAAACCA CTTGCGATAG	4860
35	GCGTCGCAAT ACAAAAACT AATCCGATAC CTATCAAGAT TTTAAATCT AATGGGAAAA	4920
	TTTTACGCAT TGTTTCTATA TCAAATGCAA TCGTAATGAT AACAAAGTGAA CTTGCGAATA	4980
	ATAATCCGCC GACGAAACCG CCACCAGGTG TATAATGTCC TGCTAAGAAA AGTGAAAAAC	5040
40	CAAAGACCAT TACCATGAAA AAGATAATAA CTGCAGCAA TTGCAAAAT AGATCATTTT	5100
	GTGTCTATT CATGATTTT CACCTCGTTA CCTGCGTTT GACGCTTTT ACGTAATTTA	5160
	ATCATTGTAT ATACAGCTAA TCCTGCGATA CCAAGCACAG ATGACTCGAA TAAAGTATCC	5220
45	ATACCACGGA AATCAACAAG TATGACGTTT ACCATGTTTT TACCGTGAGC tAAATCATAA	5280
	ACGTGCTCTT GATAAACTT AGATATCGAT TCAAAATGTC TATTTCCGTA TGCAATTAAA	5340
50	CCGATAATAA TGACGGACAA ACCAACACCA CCAGCAATTA AAGCATTAGT AAGCTGGAAT	5400
	GAGCGCTTTT CATTATAACG ATTTAAATTT GGTAAGTGGT AGAAGCATAA TAAGAACAAT	5460

55

	ATAACAATA CAGACACAGC ATATCCAAC TACTTAACA TAATGATGCT AAATAATCTT	5580
	GATTTAGCGA AAAGAATTAA AAAGGCAGCA CTTAATAATA AAATTACGAT ACAAACCTCG	5640
5	AAAATTCTAA TCGGACTAAC GTCTTTAAAA TTAATGTTGA AAGGTACTGA GAATATAGTG	5700
	ACAAATGTGA ATAAAATTAA TGCACCAAAA ATGATAACTA AATTATTACG TGAATAATCG	5760
	GTAACATAGC TATTCGTCAT CTTTTCAGAG TAGTTTGGA TAACATTTGC ACTTCTGTTG	5820
10	TACCAATAAT TGAATGTTAG TTTACCAGGT TGTCGTTGCA ACAATTTTAC CCAATAACTA	5880
	AATGTCACAA TTAGTAAGAT ACCTAAAATA TAAATCACTA ATGTTGATAA AAAGGCAGGC	5940
15	GTTAATCCAT GGAACATATG GAATTCACAA TCATCAATTA CCGTATGATT AATCGAAGag	6000
	TnAGCTGGTT CAATAATCGA ATTAGTTAAA ATGCCAGGGA ATAAACCAAA TACAATTACT	6060
	AATGTAGCTA AAATAGCTGG TGATAAAAGC ATTAATATTG ATACTTCGTG TGCTTTTTTA	6120
20	GGTAATTGTT CAGGTTTATA TTGTCCGAAA AATATATGCA TTATAAATTT AATTGAATAT	6180
	ACAAATGTGA AGACACTGCC CACTATACCA ATGATTGGGA ATAGGTAGCC TAATGTATCA	6240
	ACACTGAATA AATTTGCTTG GCTTGCTGTA AATGTTGTTT CTAAAAATGA TTCTTTTGAT	6300
25	AAGAAACCAT TGAACGGTGG TACACCAGCg CATACTTAAT GCTGTAATAA CAGTGATTGT	6360
	AAATGAAATA GGCATAATTG TTAGTAAGCC ACCTAATTTT TTAACATCAC GTGTACCAGT	6420
	AGAATGATCC ACTGCACCTG TAATCATAAA TAGGGCACCT TTAAATGTTG CATGGTTGAT	6480
30	TAAATGGAAT ATTGCAGCCG TAAATGCAGC AGCATATATT TTGCTATCAT CGCCTTGATA	6540
	GTGATAACTA ATGGCACC GA TTCCAAGCAT CGCCATAATC ATACCTAATT GGGATACTGT	6600
	TGAAATGCC AGTATACCTT TCAAGTCTTG TTGTTTGTG GCGTTTAGCG AAgCCCAGAA	6660
35	TAATGTAATT AAACCAACGA GTGTGACAGT CCATACCCAA CCTTGCGATG CTGCGAAGAT	6720
	TGGTGTTCATT CGAGCGATTA AATATAACCC TGCTTTAACC ATTGTTGCTG AATGAAGATA	6780
40	AGCACTGACT GGTGTAGGTG CTTCCATTGC ATCTGGTAGC CAAATATAAA ATGGAAACTG	6840
	AGCAGATTTT GTAAAAGCAC CAATCATGAT TAAAATCATC GCAAAAATGA AGAATGGGCT	6900
	ATTTTGAATT TCAGAAGCAT GTTGAATCAT GTACTGAATG CTAAATGATT GTGTTGGTAT	6960
45	AGCGAGTAAG ATGATACCAC CTAATAATGA TAGACCACCA AATACTGTGA TTATGAGCGA	7020
	TTTTTGAGCA CCATATATAG ATGCTTGTCG TTCGCGCCAG AATGAAATAA GTAAAAAACT	7080
	AGAAATGAC GTTAGCTCCC AGAATAAATA TAGAATAATA ACATTATCTG AAAGTACGAC	7140
50	ACCTAACATT GCACCCATAA ATAGTAATAA ATAACAATAA AAATTCCCTA GTTGTCTGTA	7200
	CTTACTTAAG TAGCCGATTG AATATAATAC TACTAACTG CCGATTCTG AAATAAGCAA	7260

55

EP 0 786 519 A2

CCAATTTAAG GTTTTCATTA CAGTATTACC TGACATCGTC GTTTTAATTA ATGTAAGCAT 7380  
ATAAATAAAT ATGACGATAG GGACAGGTAA TACGAACCAT CCTAAATGTA TACGTTTAAA 7440  
5 AAATCTATAC AGGATAGGAA TAATGAGTGC GAATATTAAC GGTAATATCA CCGCAATATG 7500  
TAACAAACTC ACTATGTTGT CCTCCTTTAA AAAATATTTA TGTATTTCAT TATACATGAA 7560  
TGATATAGTT CTGAAAAACG TACACACTCC TTGTTGTGCT TTATTTTCAG AaGTATTTAA 7620  
10 ATAAGAAGAA ACACGTCATT TTTTATTTAA AATTTTCTTT GTATTGAAGT GAATAATCTT 7680  
CTTTTAAGCG TGCTAAACTA GCTAAAGACA TTTCAGCATG TTTTGTTCG TGAGCTTTAA 7740  
GTTTAGTTTC TAAATCTGTA ATTGCTTGTT GAACTGAATC TTCATAGCGC AATACATCAA 7800  
15 CATTGAAGTC GCGTAATTGT GAACGTTTCG TATAGCGTTT TTCAAATGG CTTAATGCTT 7860  
TGCGGTCATG GAAAAATACA CCTTCAGTTT CAGTAGGGTT ATGTAAATCA CCTTGTTTCG 7920  
GGTGTTTGAT AACTTGTTCA ACTTTAACAA GGACATCGTC TCCATTTTCT TCAACAATCG 7980  
20 TGACACCATA GCTACCTGTT TTGTGTGAAA ATCGATATAG CTTCATGCTA TTTTCCTCCC 8040  
TTAAAAGTAT GTTAATATAT ATGTATCATA ACATGAATGG AGAATATAAA TGGCTAACTA 8100  
25 TCCACAGTTA AACAAAGAAG TACAACAAGG TGAATCAAA GTGGTTATGC ACACAAATAA 8160  
AGGTGACATG ACATTCAAAT TATTTCCAAA TATTGCACCA AAAACAGTTG AAAATTTTGT 8220  
GACACATGCA AAAAATGGTT ATTATGATGG AATCACATTC CACCGTGTCA TTAATGACTT 8280  
30 CATGATTCAA GGTGGCGATC CAACAGCTAC TGGTATGGGT GGCGAAAGTA TTTATGGCGG 8340  
TGCTTTTGAA GATGAATTTT CATTAAATGC ATTAACTTA TATGGCGCAT TATCAATGGC 8400  
TAACTCAGGA CCTAATACTA ATGGTTCACA ATTTTTCATT GTTCAAATGA AAGAAGTACC 8460  
35 TCAAAATATG TTAAGTCAAC TTGCAGATGG TGGCTGGCCT CAACCAATCG TTGATGCATA 8520  
TGGCGAAAAG GGTGGTACAC CATGGTTAGA TCAAAAACAT ACAGTATTCG GTCAAATCAT 8580  
TGATGGTGAA ACTACATTAG AAGATATTGC AAATACAAA GTGGGACCAC AAGATAAACC 8640  
40 ACTTCATGAT GTTGTAATTG AATCTATTGA TGTGAAGAA TAATATCTAA ACATAATTAA 8700  
CTACCAACAT TTAAACTCG GATAAAGCTA ATTTATGAAT GGATTAGTAT ATATTCCAAC 8760  
gAAAATAAAT AACTAATAT GATGAGCAAT CTCAATATAT TTATCaAGAA AGCACAGTTT 8820  
45 TTAAATAGAT GTGTATTTTA AAGATAATAG TTGAGGTTGC TTTTATGTT TTTACAGAGA 8880  
ATTGCTATTC AAATAGTAAA TAAATTGAAA ACAAAGTAGC TGGATATCAT ATTGATTTAG 8940  
50 ATAGGAATTT GTTGCTAATT TTATTTGTAA ATCCAAGTTT GTAGAATTCT TATTCATTTA 9000  
TAAAATAATA TTCGTATGAT TTGATTTTTT AATTAGTCCA CCATTTTCGAT TTGTGCTATG 9060

55

## EP 0 786 519 A2

	AACATATCAA	GGTGC GTGTA	CTGGTATTCA	ACCATACGGT	GCGTTTGTG	AGACCCCTAA	9180
	TCATACTGAA	GGACTGATT	ATATATCAGA	AATTATGGAT	GACTACGTTC	ATAATTTGAA	9240
5	GAAATTTCTA	TCAGAAGGCC	AAATTGTTAA	AGCTAAAATT	TTGTCTATAG	ATGATGAAGG	9300
	AAAGCTTAAT	CTATCATTAA	AGGATAATGA	TTACTTCAAA	AATTATGAGC	GTAAGAAGGA	9360
	AAAACAATCA	GTATTAGATG	AAATCAGAGA	AACAGAAAAA	TATGGGTTC	AAACACTTAA	9420
10	AGAACGCTTA	CCAATCTGGA	TAAAACAGTC	AAAGCGAGCA	ATTCGAAACG	ACTAAAGGAA	9480
	CAGATAAATC	GTACCGAAAA	TCATACAAAG	GGTCTGAAAT	GAAAGTTTCT	TAGACTATAA	9540
	AAGAGATTAG	TATCTATTAA	ATTTTATTAG	ATACTAATCT	CTTTTGTCT	ACGATAACGT	9600
15	AATATGaTTG	ATTCTATTTA	CACGTACAAA	TGGTTTAAGG	TGACATATCC	ATTATCTTTG	9660
	TTAGATAGAA	TCGTTGATTT	GCaATATTGT	ATGTGGATTT	GTTTTTTTTA	TTTATTTTAG	9720
20	AAATGAGAAC	TACAACTTAA	AGTATTAAAC	GAATTGCAAC	TATATAAACA	GATAATTGGA	9780
	GAATGAAAAA	ATTACATGTT	ATAGTCAACT	CAATAATTTT	AAGGAGGAAT	TAAGTAATGA	9840
	AAAGTAAATA	CGAACCATTG	TTTGATAAAG	TAGAATTACC	AAATGGAGTA	GAGTTGAGAA	9900
25	ATCGATTTGT	GTTAGCCCCT	TTAACACATA	TTTCTTCAAA	TGATGATGGT	ACTATTTTCAG	9960
	ATGTAGAACT	TCCTTATATT	GAAAAGCGTT	CACAAGATGT	TGGTATTACA	ATTAATGCTG	10020
	CGAGTAATGT	GAGTGATGTC	GGAAAAGCAT	TTCCAGGACA	GCCATCAATC	GCGCATGACA	10080
30	GTAATATTGA	AGGACTAAAA	CGATTAGCTA	CAGCAATGAA	GAAAAACGGT	GCCAAAGCAC	10140
	TCGTACAAAT	ACATCATGGC	GGTGCACAAG	CATTGCCTGA	ATTAACACCT	GATGGAGACG	10200
	TCGTAGCACC	AAGTCCAATT	TCTTTAAAAA	GTTTTGGTCA	GAAACAAGAA	CATAGTGCTA	10260
35	GAGAAATGAC	GAATGAAGAG	ATTGAACAAG	CAATCAAGGA	TTTTGGTGAA	GCAACGCGAC	10320
	GTGCAATTGA	AGCAGGGTTT	GATGGTGTG	AAATACATGG	CGCGAATCAT	TACTTAATTC	10380
40	ATCAATTTGT	ATCACCATAC	TATAATAGAA	GAAATGATGT	ATGGGCAAAT	CAATATAAAT	10440
	TCCCGGTGCG	TGTGATTGAA	GAAGTACTTA	AAGCGAAAGA	AGCGTATGGC	AATAAAGACT	10500
	TTATAGTTGG	ATACAGATT	TCTCCAGAGG	AAGCGAGTC	TCCAGGAATC	ACAATGGAAA	10560
45	TTACAGAGGA	ACTCGTTAAT	AAAATTAGCC	ATATGCCAAT	CGACTATATT	CATGTTTCAA	10620
	TGATGGATAC	GCAATGCAACG	ACACGTGAAG	GTAAATACGC	TGGACAAGAA	AGACTGCCTT	10680
	TAATTCACAA	ATGGATAAAT	GGTCGTATGC	CACTTATCGG	TATTGGTTCA	ATTTTCACAG	10740
50	CTGACGAAGC	TTTAGATGCA	GTTGAAAATG	TTGGTGTGTA	CTTAGTAGCC	ATTGGTAGAG	10800
	AGCTACTACT	GGATTATCAA	TTTGTGAAA	AAATTAAAGA	TGGACGGGAA	GATGAAATTA	10860

55



AATTTAATGA AGGGTTTTAT CCATTACCAC GTA

10953

(2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8155 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

TTTGATAnAA AACTGAATnA ATTAAATGTA TCGATTCAAC CTAATGAAGT GAATTTACAA 60  
 GTTAAAGTAG AGCCTTTTAG CAnAAAGGTT AAAGTAAATG TTAAACAGAA AGGTAGTTTA 120  
 GCAGATGATA AAGAGTTAAG TTCGATTGAT TTAGAAGATA AAGAAATTGA AATCTTCGGT 180  
 AGTCGAGATG ACTTACAAAA TATAAGCGAA GTTGATGCAG AAGTAGATTT AGATGGTATT 240  
 TCAGAATCAA CTGAAAAGAC TGTAAAAATC AATTTWCCAG AACATGTCAC TAAAGCACAA 300  
 CCAAGTGAAA CGmAGGCTTA TATAAATGTA AAATAAATAG CTAAATTAAA GGAGACTAAA 360  
 CAATGGGAAA ATATTTTGGT ACAGACGGAg TAAGAGGTGT CGCAAACCAA GAACTAACAC 420  
 CTGAATTGGC ATTTAAATTA GGAAGATACG GTGGCTATGT TCTAGCaCAT AATAAAGGTG 480  
 AAAAACACCC ACGTGTACTT GTAGGTGCGG ATACTAGAGT TTCAGGTGAA ATGTTAGAAT 540  
 CAGCATTAAT AGCTGGTTTG ATTTCaATTG GTGCAGAAGT GATGCGATTA GGTATTATTT 600  
 CAACACCAGG TGTTGCATAT TTAACACGCG ATATGGGTGC AGAGTTAGGT GTAATGATTT 660  
 CAGCCTCTCA TAATCCAGTT GCAGATAATG GTATTAAATT CTTTGGATCA GATGGTTTTA 720  
 AACTATCAGA TGAACAAGAA AATGAAATTG AAGCATTATT GGATCAAGAA AACCCAGAAT 780  
 TACCAAGACC AGTTGGCAAT GATATTGTAC ATTATTGAGA TTACTTTGAA GGGGCACAAA 840  
 AATATTTGAG CTATTTAAAA TCAACAGTAG ATGTTAACTT TGAAGGTTTG AAAATTGCTT 900  
 TAGATGGTGC AAATGGTTCA ACATCATCAC TAGCGCCATT CTTATTTGGT GACTTAGAAG 960  
 CAGATACTGA AACAATTGGA TGTAAGCCTG ATGGATATAA TATCAATGAG AAATGTGGCT 1020  
 CTACACATCC TGAAAAATTA GCTGAAAAAG TAGTTGAAAC TGAAAGTGAT TTTGGGTTAG 1080  
 CATTTGACGG CGATGGAGAC AGAATCATAG CAGTAGATGA GAATGGTCAA ATCGTTGACG 1140  
 GTGACCAAAT TATGTTTATT ATTGGTCAAG AAATGCATAA AAATCAAGAA TTGAATAATG 1200  
 ACATGATTGT TTCTACTGTT ATGAGTAATT TAGGTTTTTA CAAAGCGCTT GAACAAGAAG 1260  
 GAATTAAATC TAATAAACT AAAGTTGGCG ACAGATATGT AGTAGAAGAA ATGCGTCGCG 1320

	CTGGTGATGG TTTATTAACT GGTATTCAAT TAGCTTCTGT AATAAAAAATG ACTGGTAAAT	1440
	CACTAAGTGA ATTAGCTGGA CAAATGAAAA AATATCCACA ATCATTAAAT AACGTACGCG	1500
5	TAACAGATAA ATATCGTGTT GAAGAAAATG TTGACGTTAA AGAAGTTATG ACTAAAGTAG	1560
	AAGTAGAAAT GAATGGAGAA GGTCGAATTT TAGTAAGACC TTCTGGAACA aACCATTAGT	1620
	TCGTGTCATG GTTGAAGCAG CAACTGATGA AGATGCTGAA aGATTTGCAC AACAAATAGC	1680
10	TGATGTGGTT CAAGATAAAA TGGGATTAGA TAAATAAATA CTGTATTACA AATGAGCCGA	1740
	TGCGTATGcA nTcgTITTTT GTGTTGTAG AAATAATTTA TAGTACAAAC GTAAATGAT	1800
	ATAAACAAAA TAAAAACAAA GTAATCAATA TGTAATATAA AATACACTGG TACTCAATAT	1860
15	ATAATGATGA TAAATTTAAT TTTAATTAGA TAGAGTTGCT TTGTGTTTTT AACGCAGATG	1920
	CTACTACTTA TCTTAACAGT TGATTAAGTG AAATCATTTA ACAGCGAGAA TAATCAACCA	1980
20	GGAGGATGAC TTAATGAATT TATTCAGACA ACAAAAATTT AGTATCAGAA AATTTAATGT	2040
	CGGTATTTTT TCAGCTTTAA TTGCCACTGT TACTTTTATA TCTACTAACC CGACAACAGC	2100
	GTCTGCAGCA GAGCAAAATC AGCCTGCACA AAATCAACCA GCACAACCAG CTGATGCCAA	2160
25	TACACAGCCT AACGCAATG CTGGTGCTCA AGCTAATCCT ACAGCACAGC CAGCTGCACC	2220
	TGCCAACCAA GGACAACCAG CAGTACAACC AGCAAACCAA GGTGGACAGG CTAATCCAGC	2280
	AGGAGGAGCA GCACAACCAA ATACACAACC AGCTGGACAA GGTGATCAAG CTGATCCGAA	2340
30	TAACGCTGCA CAAGCACAACT CTGGAAATCA AGCAACACCG GCAAACCAAG CAGGTCAAGG	2400
	AAATAACCAA GCAACACCTA ATAATAATGC AACACCGGCA AATCAAACAC AGCCAGCGAA	2460
	TGCTCCAGCA GCAGCGCAAC CAGCAGCACC TGTAGCAGCA AACGCACAAA CTCAAGATCC	2520
35	AAATGCTAGC AATACTGGTG AAGGCAGTAT TAATACGACA TTAACATTTG ATGATCCTGC	2580
	CATATCAACA GATGAGAATA GACAGGATCC AACTGTAAGT GTTACAGATA AAGTAAATGG	2640
40	TTATTCATTA ATTAACAACG GTAAGATTGG TTTCGTTAAC TCAGAATTAA GACGAAGCGA	2700
	TATGTTTGAT AAGAATAACC CTCAAAATA TCAAGCTAAA GGAAACGTGG CTGCATTAGG	2760
	TCGTGTGAAT GCAAATGATT CTACAGATCA TGCTAACTTT AACGGTATTT CAAAACTGT	2820
45	AAATGTAAAA CCAGATTGAG AATTAATTAT TAACTTTACT ACTATGCAAA CGAATAGTAA	2880
	GCAAGGTGCA ACAAATTTAG TTATTAAAGA TGCTAAGAAA AATACTGAAT TAGCAACTGT	2940
	AAATGTTGCT AAGACTGGTA CTGCACATTT ATTTAAAGTA CCAACTGATG CTGATCGTTT	3000
50	AGATTTACAA TTTATTCCTG ACAATACAGC AGTTGCTGAT GCTTCAAGAA TTACAACAAA	3060
	TAAAGATGGT TATAAATACT ATTCATTCAT TGATAATGTA GGTCTATTCT CAGGATCACA	3120

55

EP 0 786 519 A2

	TAATACTGAA ATCGGTAACA ATGGTAATTT TGGTGCTTCA TTAAGAGCAG ATCAATTTAA	3240
	ATATGAAGTA ACATTACCAC AAGGTGTAAC TTACGTTAAT AATTCATTAA CTACAACATT	3300
5	CCCTAATGGT AATGAAGACA GTACAGTATT GAAAAATATG ACTGTTAATT ATGATCAAAA	3360
	TGCAAATAAA GTTACATTTA CAAGCCAAGG TGTGACAACG GCACGTGGTA CACACACTAA	3420
	AGAAGTTTTA TTCCCAGATA AATCTTTAAA ATTATCATAT AAAGTTAATG TTGCGAATAT	3480
10	CGATACACCT AAAAATATTG ATTTTAATGA AAAATTAAACA TATCGTACTG CTTCAGATGT	3540
	TGTAATTAAT AATGCGCAAC CAGAAGTACA CTAATGCGAG ATCCATTTTC AGTAGCGGTT	3600
	GAAATGAACA AAGATGCGTT GCAACAACAA GTAAACTCAC AAGTTGATAA TAGTCATTAC	3660
15	ACAACAGCAT CAATGCGAGA ATACAATAAA CTTAAACAAC AAGCAGATAC TATTTTAAAT	3720
	GAAGATGCGA ATCATGTTAA AACTGCAAAT CGTGCACTC AAGCGGATAT TGATGGTTTTA	3780
	GTAATAAAT TACAAGCTGC ATTAATTGAT AATCAAGCAG CAATTGCTGA ATTAGATACT	3840
20	AAAGCTCAAG AAAAGGTTAC AGCAGCACAA CAAAGTAAAA AAGTTACGCA AGATGAAGTT	3900
	GCAGCACTTG TAACTAAAAT TAACAATGAT AAAAATAATG CAATCGCAGA AATTAATAAA	3960
25	CAAACTACAG CACAAGGTGT CAAACTGAA AAAGATAATG GTATCGCAGT GTTAGAACAA	4020
	GATGTGATTA CACCAACAGT TAAACCTCAA GCGAAACAAG ATATTATCCA AGCAGTTACA	4080
	ACTCGTAAAC AACAAATTAA AAAGTCAAAT GCATCATTAC AAGATGAAAA AGATGTAGCA	4140
30	AATGATAAAA TTGGTAAAAT TGAAACAAAG GCAATTAAAG ATATTGATGC AGCAACAACA	4200
	AATGCACAAG TAGAAGCCAT TAAAACAAAA GCAATCAATG ATATTAATCA AACTACACCT	4260
	GCTACAACAG CTAAAGCAGC AGCTCTTGAA GAATTGACG AAGTTGTTCA AGCACAAATT	4320
35	GATCAAGCAC CTTTAAATCC TGATACAACA AATGAAGAAG TAGCGGAAGC TATTGAACGT	4380
	ATTAATGCAG CTAAAGTTTC TGGTGTTAAA GCAATTGAAG CGACAACGAC TGCACAAGAT	4440
	TTAGAAAGAG TTAAAAACGA AGAAATCTCA AAAATTGAAA ATATTACTGA CTCTACGCAA	4500
40	ACAAAAATGG ATGCCTATAA TGAAGTTAAA CAAGCTGCAA CAGCTAGAAA AGCTCAAAAT	4560
	GCTACAGTTT CAAATGCAAC AAATGAAGAA GTAGCAGAAG CTGATGCAGC AGTAGATGCA	4620
	GCTCAAAAGC AAGGTTTACA TGACATCCAA GTTGTTAAAT CAAAACAGGA AGTTGCTGAT	4680
45	ACAAATCAA AAGTATTAGA TAAATCAAT GCAATTCAAA CACAAGCAAA AGTTAAACCT	4740
	GCAGCTGATA CGGAAGTAGA AAACGCATAT AATACACGTA AACAAGAAAT TCAAAATAGC	4800
50	AATGCTTCAA CTACAGAAGA AAAACAAGCT GCATATACAG AATTAGATAC TAAAAAGCAA	4860
	GAAGCAAGAA CAAATCTTGA TGCTGCAAAT ACAAACAGTG ATGTAACAAC AGCTAAAGAC	4920

55

	GCGGAAATCG	CTCAAAAAGC	AAGTGAACGT	AAAACAGCAA	TTGAAGCAAT	GAATGATTCTG	5040
	ACTACTGAAG	AACAACAAGC	AGCGAAAGAC	AAAGTGGATC	AAGCAGTAGT	TACTGCAAAC	5100
5	GCTGATATAG	ATAATGCTGC	AGCAAACAAT	GATGTGGATA	ATGCAAAAAC	TACAAATGAA	5160
	GCTACAATCG	CAGCCATTAC	ACCTGATGCA	AATGTTAAAC	CAGCAGCAAA	ACAAGCAATT	5220
	GCAGATAAAG	TACAAGCTCA	AGAAACAGCA	ATTGATGGAA	ATAACGGCTC	AACAACCTGAA	5280
10	GAAAAAGCAG	CTGCTAAACA	ACAAGTTCAA	ACTGAAAAAA	CAACAGCTGA	TGCCGCAATA	5340
	GATGCAGCAC	ATACAAATGC	GGAAGTTGAA	GCGGCTAAAA	AAGCAGCAAT	TGCTAAAAAT	5400
	GAAGCGATTG	AGCCAGCAAC	AACAACATAA	GATAATGCGA	AAGAAGCAAT	TGCTACGAAA	5460
15	GCGAATGAAC	GTA AACAGC	AATCGCTCAA	ACGCAAGACA	TTACTGCTGA	AGAAATTGCA	5520
	GCGGCTAATG	CGGACGTAGA	TAATGCTGTG	ACACAAGCAA	ATAGCAACAT	TGAAGCTGCT	5580
20	AATAGTCAAA	ATGATGTAGA	CCAAGCGAAA	ACGACAGGTG	AAAATAGTAT	TGATCAAGTA	5640
	ACACCAACAG	TTAATAAAAA	AGCAACTGCA	CGTAATGAAA	TCACAGCAAT	TTTAAATAAC	5700
	AAATTGCAAG	AGATTCAAGC	TACGCCAGAT	GCAACAGATG	AAGAAAAACA	AGCAGCTGAT	5760
25	GCTGAAGCAA	ATACTGAAAA	TGGTAAAGCA	AATCAAGCCA	TTTCAGCAGC	AACTACTAAC	5820
	GCACAAGTTG	ATGAAGCTAA	AGCAAATGCA	GAAGCAGCGA	TTAATGCGGT	AACACCAAAA	5880
	GTTGTGAAGA	AACAAGCGGC	TAAAGATGAA	ATTGATCAAT	TACAAGCAAC	GCAAACAAAT	5940
30	GTTATCAATA	ATGATCAGAA	CGCTACAACA	GAAGAAAAAG	AAGCAGCTAT	TCAACAATTA	6000
	GCAACAGCAG	TTACAGACGC	GAAAAATAAT	ATTACAGCTG	CAACTGATGA	TAATGGTGTA	6060
	GATCAGGCGA	AAGACGCTGG	AAAGAATTCA	ATTCAAAGCA	CGCAACCAGC	AACAGCGGTT	6120
35	AAATCAAATG	CTAAAAATGA	TGTTGATCAA	GCTGTGACAA	CTCAAAATCA	AGCAATTGAT	6180
	AATAEAACTG	GTGCTACAAC	TGAAGAGAAA	AATGCAGCAA	AAGATTTAGT	TTTAAAAGCT	6240
	AAAGAAAAAG	CGTATCAAGA	TATCTTAAAT	GCACAAACAA	CTAATGATGT	TACGCAAATT	6300
40	AAAGATCAAG	CAGTTGCTGA	TATTCAAGGT	ATTACTGCAG	ATACAACAAT	TAAAGATGTT	6360
	GCGAAAGATG	AATTAGCAAC	AAAAGCAAAC	GAACAAAAAG	CGCTTATTGC	ACAAACTGCA	6420
45	GATGCGACTA	CTGAAGAAAA	AGAACAAGCA	AATCAACAAG	TAGACGCACA	ATTAACACAA	6480
	GGTAATCAAA	ATATTGAAAA	TGCACAGTCA	ATCGATGATG	TAAACACTGC	AAAAGATAAT	6540
	GCAATTCAAG	CAATTGACCC	AATTCAAGCA	TCAACAGATG	TTAAAACGAA	TGCAAGAGCG	6600
50	GAATTGCTAA	CTGAAATGCA	AAATAAAATA	ACTGAAATAC	TTAATAATAA	TGAGACTACT	6660
	AATGAAGAAA	AAGGTAACGA	TATTGGACCA	GTTAGAGCAG	CATATGAAGA	AGGTTTAAAT	6720

55

AAAGTTCAAC AACTTCATGC AAATCCTGTT AAGAAACCAG CAGGTAAAAA AGAATTAGAT 6840  
 CAAGCTGCAG CTGATAAGAA AACACAAATA GAACAAACAC CAAATGCATC ACAACAAGAA 6900  
 5 ATTAATGATG CAAAACAAGA AGTTGATACT GAATTAAATC AAGCGAAAAC AAATGTCGAT 6960  
 CAATCATCAA CAAATGAATA TGTGATAAT GCAGTTAAAG AAGGAAAAGC TAAAATTAAT 7020  
 GCAGTTAAAA CATTTAGTGA GTACAAAAAA GATGCTTTAG CTAAAATTGA AGATGCATAT 7080  
 10 AATGCTAAAG TAAACGAAGC GGATAACTCT AACGCATCGA CTTCAAGTGA AATTGCTGAA 7140  
 GCGAAACAAA AACTTGCTGA ATTAAAACAA ACTGCGGATC AAAATGTTAA TCAAGCTACT 7200  
 TCTAAAGATG ACATTGAAGT TCAAATTCAT AATGACTTAG ATAATATTAA CGATTACACA 7260  
 15 ATTCCAACAG GTAAAAAAGA ATCAGCTACA ACAGATTTAT ATGCTTATGC AGATCAGAAG 7320  
 AAAAATAATA TTTCAGCTGA CACTAATGCA ACACAAGATG AAAAGCAACA AGCAATTAAG 7380  
 CAAGTTGACC AAAATGTTCA AACTGCATTA GAAAGCATTG ATAATGGTGT GGATAATGGT 7440  
 20 GACGTTGATG ATGCATTAAAC ACAAGGTAAA GCAGCAATTG ATGCTATTCA AGTAGATGCT 7500  
 ACTGTTAAAC CTAAAGCGAA CCAAGCTATT GAAGTTAAAG CAGAAGATAC GAAAGAATCT 7560  
 25 ATTGATCAAA GTGACCAGTT AACTGCTGAA GAAAAAATG AAGCATTAGC AATGATTAAA 7620  
 CAAATTACAG ATCAAGCTAA ACAAGGTATT ACTGATGCAA CAACAACTGC TGAAGTTGAA 7680  
 AAAGCGAAAg cTCaAGGACT TGAAGCATTT GATAACATTC AAATCGACTC AACAGAAAAA 7740  
 30 CAAAAGCTA TCGAAGAATT AGAAACTGCA CTAGACCAGA TTGAAGCAGG TGTAAATGTC 7800  
 AACGCTGATG CTACAACTGA AGAAAAAGAA GCGTTTACGA ATGCTTTAGA AGACATTTTA 7860  
 TCAAAGCAA CTGaAGATAT TTCTGATCAA ACTACAAATG CAGAAATCGC TACTGTCAAA 7920  
 35 AATAGTGC GC TTGAACAACT TAAAGCACAA CGTATTAATC CTGAAGTTAA GAAAAATGCT 7980  
 TTGGAAGCAA TCAGAGAAGT GGTTAACAAG CAAATAGGAA tAATTAAAAA TGCAGATGCA 8040  
 GATGCATCGG CGGAAAGAnA TTGCACGTAC GGGATTTAGG TAGATATTTT GGACCGATTT 8100  
 40 GCTGGATAAA TTTAGGGTnA AACCCCAACC AATGCCGAAG TTGCCTGAAT TACCA 8155

(2) INFORMATION FOR SEQ ID NO: 64:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1630 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

55

EP 0 786 519 A2

	CTGTTTTATT TGCAGCACCC ATACTGGAAA TCACTTTAAT CCCTCGGTCA AGACACTCTT	120
	TCATTAAGTG TACTTTGTAC ATTATTGTAT CACTTGCATC TACAAAATAA TCTATATCGT	180
5	AGTTATCGAA AATTTCTTCA TATGTCTCTT CTGTATAAAA CATATGTAAG GCGTGACTT	240
	TACAATCTGG ATTAATTAAT TTAATACGTT CTTCCATCAA AGAAACTTTA CTTTGTCTTA	300
	CCGTTGTAGT TAAAGCGTGT AATTGTCTGT TTACATTTGT AATATCAACA TCATCTTTAT	360
10	CTATTAATAT AATATGACCA ATATTCGTTT TTGCTAATGC TTCAGCAGCA AATGAACCAA	420
	CACCTCCAAC GCCAAGTATG ACAACAGTTT GTTGCTTCAA TAAATCTAAA CCTTGTGTGC	480
	CAATCGCTAG TTCATTTCTT GAAAATTGAT GTTTCATTAT TTTACCTCTT TCACTGATTT	540
15	ATACATAAGT ACATAGTAAC TTAATAATTT ATATTTAGCA TTATCACTTT GATTATTTTC	600
	CCAAAATTCA ACGAGGAAAC ATTTATTAAA CGCTATAAAA CCCAACTAAT TCTTTATTAA	660
	AAACTTAAAG AAACGCATAA AAATACGCAA GACAAAGTCT TCGGTATCGA TAGAGTCCGT	720
20	ATTGCCGTAG TTATAATAGC TTGATCATTC GGCCTGTTAT ATACAGGTGG GTGCCCTGTT	780
	TCTTGTTTTG TACGTCCTTC ATATAAGGCG TGTACGCTGC AAGAAAACCC ATTGGGCTCC	840
25	CTTGATCAAA GAGTGTTAGG CCCAAATTAA AAAGCAAAC TACGAACAAC TCAGATGACT	900
	ATCTTATGAT GTTATATTAC CACATAATTA AAATTAATGA AATTATAACA AACCAAAGTT	960
	TATTGATTTT TTAATAATTA GTGACGAATT CGCAAAGAAA GTTCTTCTAA TTGTTTATCA	1020
30	GAAACTTCAC TAGGCGCATT CGTTAATAAA CATGTAGCAG ATGCTGTTTT AGGGAATGCG	1080
	ATTGTATCTC TCAAGTTTGT TCTATTAGTC AATAACATGA CTAATCGGTC TAATCCTAAT	1140
	GCAATACCGC CATGTGGTGG TGCACCATAT TTAAATGCAT CTAGTAAAGAA GCCGAACGTG	1200
35	TCCTGTGCTT GTTCTTTAGT AAATCCAAGA ACTTCGAACA TTTTCTCTTG TAACTACCA	1260
	TCATGAATTC TGATTGAACC GCCACCTAAT TCATAACCAT TTAATACTAT GTCATAAGCA	1320
	TTTGCTCAG CTTCTCTGG CGCAGTGCCA AGCTTAGCAA TATCAGCTTC TTTTGGAGAT	1380
40	GTAAATGGAT GATGTGCTGC AACGTAACGT TTCGCATCTT CATCATATTC TAATAATGGC	1440
	CAATCTGTCA CCCATAAGAA GTTAAATTTT GTTTCATCGA TTAAACCTAA TTCTTTAGCT	1500
	AATTTGACAC GTAATGCACC TAACTTTTGT GCAACGACAT TTGGTCTGTC TGCAACAAAC	1560
45	ATTACTAAGT CACCAGCTTC AGCACCAGTT AATGTAAGTA ATGTTTCAAC ATTTTCTGTT	1620
	cAAAGAAACG	1630

50 (2) INFORMATION FOR SEQ ID NO: 65:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 732 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

	CAATTGGACA TCTTGATGA AAAGGACAAC CTTGCGGCGG ATTACTTGGC GAAGGTAATT	60
10	CTCCTTTTAA TATAATTCTA TTGTTATTAT GTTTATCAAT TTGTGGTATT GATGAAATCA	120
	ACGCTTTTGT ATATGGATGT TTGGGATTTT CATAAATTC TTTATCAGAT GCGATTTCAA	180
	CTATATGACC TAAATACATA ACTCCAATGA CATCACTTAT ATGTTTTACT ACACTTAAAT	240
15	CATGTGCGAT AAATAAATAG CTTAAGTTAA ATTGTTCTTG TAAATCTTTT AATAAATTCA	300
	GTACTTGAGA TTGAACAGAT ACATCTAATG CACTTACAGG CTCATCAGCA ACAATTAAAC	360
	TCGGACGCAA AGCCAATGCT CTTGCAATTC CCACTCTTTG TCTCTGTCCA CCTGAAAATT	420
20	CATGTGCATA TTATAATAT GCATCTTCAC TTAGGCCAAC ACATTTTAAT AAATATAGTA	480
	CTTCTTTTTT TATTCTTCT TTTGGCAATT TTTTATAATT TAAAATAGGT TCTGAAATGA	540
25	TATCTCCAAC CATTTGCATC GGATTCAATG ATGCATACGG ATCTTGAAAT ATCATCTGAT	600
	ATTGTTGTCG TGATTTTCTG AGTTTTTTAC CTTGTAATCT TGTTATATCT TCACCATTAA	660
	CAATTATTGA GCCTGAAGTT GCATCTTCAA GCCTGATAAT CACTTTACCT AACGTTGACT	720
30	TACCACAACC CG	732

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:

35	(A) LENGTH: 5838 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

	AATATATTCA TATGTTTCAT CAACAATATT AGCTGCTTTT TGAATTAAAG CAATTCGTC	60
	AGCATCTTTG ACGTCTCTAA TTTTATCTAC AGTATTAGAA ATGCTTATTA ATGATATACG	120
45	GCTTTTATTT AATTCAAGGT ATGTATCATA ACTTACATGA TGCCCCTCAA AACCTACATT	180
	TTCAAAATTT TCTTGGTGTA GCAATTCTTT AATCTCACCA ATAATAGTAG ATTTACGATT	240
50	AATAATTTC TAATTTGGCG CCTGCTTAGT TGCTTGATCA ATATATCTAA AGTCTGTTAT	300
	CAAATATTGT TTATCTTTAG ATATGATAAG TGCTCCACTG GTACCAGTAA AACCTGATAA	360
	ATATCTTCTA TTGTAATCCG AAAGAATGAT AATCGCATCT AAATGTTTTT GTTCTAAAAA	420

55

	CAACTTTATA CATTAAAATA ATATCATAAT AAGGATAAAA AATAATAGAT ATTGATTTTA	540
	GGGAGATAGT AATGAAAAAA TTGGTTTCAA TTGTTGGCGC AACATTATTG TTAGCTGGAT	600
5	GTGGATCACA AAATTTAGCA CCATTAGAAG AAAAAACAAC AGATTTAAGA GAAGATAATC	660
	ATCAACTCAA ACTAGATATT CAAGAACTTA ATCAACAAAT TAGTGATTCT AAATCTAAAA	720
	TTAAAGGGCT TGAAGAGGAT AAAGAAAACA GTAAAAAAC TGCATCTAAT AATACGAAAA	780
10	TTAAATTGAT GAATGTTACA TCAACATACT ACGACAAAGT TGCTAAAGCT TTGAAATCCT	840
	ATAACGATAT TGAGAAAGAT GTAAGTAAAA ACAAGGCGA TAAGAATGTT CAATCGAAAT	900
	TAAATCAAAT TTCTAATGAT ATTCAAAGTG CTCACACTTC ATACAAAGAT GCTATCGATG	960
15	GTTTATCACT TAGTGATGAT GATAAAAAAA CGTCTAAAA TATCGATAAA TTAACTCTG	1020
	ATTTGAATCA TGCATTGAT GATATTAAAA ATGGCTATCA AAATAAGAT AAAAAACAAC	1080
20	TTACAAAAGG ACAACAAGCG TTGTCAAAT TAACTTAA TGCAAATCA TGATAGGAGT	1140
	CTTTTAATGC GTAATATAAT ATTTTATCTT GTACTTATTA TTGCTGCGAT TGGATTAGTA	1200
	ATGAATCTAG ATGCCTTTAT TTTTCAATC GTCAGAATGT TAATCAGCTT TGcgTAaTAG	1260
25	CTGGTATTAT TTATCTGATT TATTATTTCT TCATCTTAAC TGAAGACCAA CGCAAATATC	1320
	GCAAAGCAAT GCgTrAaGTA TAAAAGAAAT CAAAGAAGAA AATAGATAAA AAAACGGAAG	1380
	CACTTGTAGG TAAAATAGTC TACGTGCTTC CATTTTTTAT TCTAAAACT ACTTTCTAAA	1440
30	CATCCATTCA TCTGAACGAT ATTTTTCAGT TAATTCTTCC ACTTCTGCCA ATTGAGCTTC	1500
	TGtTAATTCA AGTGGCTTTA ATTCTATATT TAAACCTTTC TTAAAACCTT TCTCGAAAGC	1560
	TTCTTCCATT TGACTAATAG TAATGTGTTT ATCTGAAATA TCATTGATGG CAACTGCTTT	1620
35	TTCAACGAAT GCCTCTTCA TTTTAATTT TAATCTTTCA TTTTATAAA TrAACATATC	1680
	AAACgTTCA TCAATATCAA TATCTTGTA AATCGAACCG TGTTGGAGGA TTACGCCCTT	1740
40	TTGTCTCGTT TGAGCACTCC CAGCAATCTT ACGGCCTTCA ACAACTAGCT CATACCAACT	1800
	TGGTGCATCA AAACACACTG AACTTCGAGG TTGTTTTAAT TTTTGACGCT CTTCAGGCGT	1860
	TTTAGGTACC GCAAAATAAG TATCAAATCC TAAGTTTTTA AATCCTTCTA ATAATCCTTG	1920
45	TGAAATCACT CTGTACGCTT CTGTAAGTGT AGAAGGCATA TTCGGATGCG ATTCAGGCAC	1980
	AATCACACTG TAAGTTAACT CTTTATCATG TAGCACCCCA CGGCCACCAG TTTGACGCCT	2040
	TACGAGACCA AAACCTTTCT CTTTAACCTT ATCAATATCA ATTTCTTTT GTAGCCTTTG	2100
50	GAAATACCCT ATTGATAATG TTGCAGGATT CCATGTGTAA AAACGTATAA CTGGATCAAT	2160
	TTCACCTCTA GAGACAAAAT TTAATAACGC TTCATCCATT GCCATATTAT AATATGGGTC	2220
55		



	AAATGTATAA TATTTGATTC GCTAATTAAT CAATTTAACT AAATGAATAA TAATTGCAAT	2340
	TCTTTAGTGA AATATTTTGA TAATTTGACC TAACAGTCTT ATAATTATAT TATCGTTTAA	2400
5	TTAGGGAGGA TGCAAGATGA GTGCTAGTTT GTACATCGCA ATAATTTTAG TTATAGCAAT	2460
	TATTGCTTAT ATGATTGTTC AACAAATTCT TAACAAGCGA GCTGTTAAAG AATTAGATCA	2520
	AAATGAATTC CATAATGGGA TTAGAAAAGC TCAAGTCATC GATGTTAGAG AGAAAGTTGA	2580
10	CTATGACTAC GGTCACTTA ATGGGTCTCG CAATATTCCT ATGACAATGT TCAGGCAACG	2640
	ATTCCAAGGA TTAAGAAAAG ATCAACCGGT ATACTTATGT GATGCCAATG GGATTGCTAG	2700
15	CTATAGAGCC GCTCGTATTT TGAAAAAGAA TGGATATACA GATATCTATA TGTTAAAAGG	2760
	CGGCTATAAA AAATGGACTG GAAAAATAAA GTCTAAAAAA TAGTTTTTGT AAATTTAATA	2820
	TACGATTTAA TAAAATCTGA GTGTTAATTG ATCATCAATA ACAATACTCA GATTTTAATT	2880
20	TTTTAACAAA GTCTGTACT ATATTTCTCT AGCTTCACTG ATCATTAAAC TTAGTTTCAG	2940
	CATAATAAAG AAAGTTCAGC TCATTTTCAA TACGATTCAA TTACCGCAAT CTAAAAAATG	3000
	AAAAGACAAT TTCTATGAAA GAATAATACC AAACCCTAAG AGTTATTACT TCGGTTTAGT	3060
25	TTTCTGTTT AAATAGAAAT TGTCTTTTTC AATTGATTTT GAAACCATTA TCCTTAAATC	3120
	TTCATACAAA GTTAGAATAA TAATCTCGG AATATGTGTT TAATACTTTA TTTTTCCTGT	3180
	TTAAGATTTT CAAACTTTAA TATTGGTTTA CGAGCAGCTG TAGCTTCGTC TAATCGATCA	3240
30	ATCACAGTTG TATGTGGTGC TTCTAGCacT TTATCAGGAT CATTTTTAGC TTCTTCAGCA	3300
	ATACTAATTA ATGTATCGAT AAAATAATCA AGTGTTTCTT TAGACTCTGT CTCAGTCGGT	3360
35	TCAATCATCA TACCTTCTTC AACATTTAAT GGGAAGTATA TTGTTGGTGG ATGTACACCG	3420
	AAATCTAATA ATCGCTTAGC CATGTCTAAA GTACGTACAC CAAATTCTTT TTGACGCACA	3480
	CCACTTAACA CAACTCGTG TTTACAATAT TGTTTATAAG GTATTTCAA GTGTTTAGAT	3540
40	AAACGTGCTT TAATATAATT CGCATTAAGA ACCGCTGCTT CAGAAACCTC TTAAAGTCCA	3600
	GTTGCTCCCA TAGTTCGAAT ATACGTATAA GCTCTTAAGT AAATACCAA GTTACCATAA	3660
	AATGGTTTTA CACGTCCGAT AGAATTTTTA ATGTCATTAT CATATTTAAA TTTGTCGCCA	3720
45	TCTTTAATAA CCATTGGCTT TGGTAAGTAA CTGCTAGTT CTTTTACTAC ACCGACTGGA	3780
	CCTGAACCAG GACCGCCACC ACCATGTGGA CCAGTAAATG TTTTATGCAA GTTTAAATGA	3840
	ACAGCATCAA ATCCCATATC TCCTGGGCGA ACTTTGTCCA TAATAGCGTT TAAATTCGCA	3900
50	CCATCATAAT ATAATAGACC ACCAGCATT TGGACGATTT CACGGATTTT CATAATATTT	3960
	TTTTCGAAAA TACCTAAAGT GTTTGGATTA GTTAACATAA TAGCTGCTGT ATTTTCATTT	4020
55		

EP 0 786 519 A2

	GATTTAAATC CTGCAAATGa AGCTGAGGCT GGaTTCGTAC CATGCGCAGA ATCTGGcACA	4140
	ATGACTTCAT CACGATGACC TTCACCATTA TTCTCATGGT AAGCTTTAAA TATCATCAAT	4200
5	GCAGTCCATT CACCATGTGC GCCAGCAGCT GGTGTGAATG TCACCTCATC CATACCAGTA	4260
	ATTTCTTTTA ATTCTTCTTG CAAACTATAA ATAATTTCTA ATGAACCTTG AACTTGATCT	4320
	TCATCTTGTA ATGGATGTGA TTCACTAAAT CCTGGTATTC TAGCAACCTT TTCATTAATT	4380
10	TTAGGGTTAT ACTTCATCGT ACATGAACCC AATGGATAAA ATCCGTTGTC TACACCGAAA	4440
	TTTTTATTTG AAAGTTCAGT ATAATGACGT ACTAAGTCTA GTTCAGCAAC TTCAGGAAAC	4500
	TCCGCTTTGT TTTTACGAAT AAATTTATCA TCTAACAATG ACTCAACAGA ATTTGTTTTA	4560
15	ATATCACTTT TTGGTAATGA ATATGCATAT CTGCCTTCAC GAGATCTTTC AAAAATTAAT	4620
	GGACTTGATT TACTAGTCAT TTAAGTCACC AGCCTTTTCT ACAAATGTAT CGATTTCATC	4680
20	TTTTGTCTT AATTCAGTTA CAGCTATTAA CATGTGATTT TTAAAGTCGT CTGAAACAAC	4740
	ACCTAAATCA AAACCACCGA TAATATTGTA CTTCACTAAT TCCTCGTTAA CTTGTTGAAT	4800
	TGGTTTGTC AATTTGACTA CAAACTCATT GmnaAGnTGT ACCATCTAAT ACTTCAAAAC	4860
25	CTTTTTTAAT AAATTGTTGT TTAGCATAGT TAGCATGTTT TATATTTTGA ACTGCAATAT	4920
	CATAGATACC TTGTTTACCA AGTGCTGACA TTGCAATTGA TGaCGcTAAA GCATTTAATG	4980
	CTTGTTTAGA ACAAATATTA GATGTCGCTT TATCGCGTCG AATATGTTGT TCACGTGCTT	5040
30	GTAATGTTAA TACAAAGCCA CGATTACCTT CATCATCTTG TGTTTGACCG ACTAATCTAC	5100
	CTGGCACTTT ACGCATTAAC TTTTTCGTCG TTGCAAAATA TCCACAATGT GGCCCACCGA	5160
	ATTGAGCAGG AATTCCGAAT GGCTGAGTAT CACCTACAAC AATATCTGCA CCAAATGAAC	5220
35	CTGGAGGTGT AAGTAATCCC AATGCTAATG GATTTGCATA TACGATAAAT AATGCTTTTT	5280
	TATCTFCAAT AAAGCTATGA ATCTTTTCAA GATCTTCAAT TGAACCGTAA AAGTTTGGAT	5340
40	ATTGTACTGC AACAGCTGCT GTTTCATCAT CCACTGCTGC TTCTAATTTT TTCAAATCTG	5400
	TAACAGTGCC ATCTAAATCG ATTTCCACTA CTTCGAATTC CTTACGCGTC TTAGCATAAG	5460
	TATGAAGTAC TTGTAATGCT TGATAATGTA AACCTTTTGA GACTACAATT TTATTTTCT	5520
45	TTGTTTGAAT AAATGCTAAG ATACATGCTT CAGCAAAGCT AGTCATCCA TCATACATAG	5580
	AAGAATTTGC TACATCCATA TCTGTTAATT CACAAATTAA AGTTTGAAC TCAAAAATGG	5640
	CTTGTAATTC ACCTTGAGAA ATTTCCGGTT GATATGGCGT ATATGCTGTG TAAAATCTG	5700
50	ATCTTGAAAT CATAGCATCC ACAAATGATG GCGCGTAATG ATCATAAACA CCAGCACCCA	5760
	rAAATGATGT ATGCGTTTCT TTAGTGATAT tCTTGCTkGC AATGGGGATT TAAACnTCTA	5820

55

## (2) INFORMATION FOR SEQ ID NO: 67:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18355 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

5 ATrATAATTG GCTTTGCTAA TAATTACTTC CCTGAATTAC aAGTATTAGC AAACGAAATA 60  
 15 AAATCTGATA TGGCTAGTTC ATTAAAACAA TGATATTTTT ATTTAAATTT TTaAAGCTTT 120  
 GTACGAAATT GTACAAAGCT TTTTGGTGC GTATTGTATG GGCAACAAC TACGATGAA 180  
 AATCCGTTAC AGGATTGGTA ATAGGAAATG TTAGCGAAAG ACAAGGGTAT CCATTGTAGA 240  
 20 TTAACAAAAG GACGTTTCCA CAAGTGTGGG TTATTCTCAC TAAAGCAATA CGCAGAGACA 300  
 ACTTACGTAA AATTTTGAAC TGACTAGAAC GGAACCTCTA CTCAATTATT GATAAAAATT 360  
 TTCAAAAAGA CTTGAATGTG CTGAGAATAC GAAGTTTATG GAAGGATTAT CAAAATATAA 420  
 25 ATGTGCATTC ATTACAACC TTTATTGACA ATGATTCTCA ACTAATATAG TATATAATCA 480  
 AATCGTAATA GTTACGATTT GTTTTCTGCA ACTTTTTTGA AGTTTTAGTT GAGGTGAAAA 540  
 CAATAAAAGC ATCTAAGTGA ATGTAGTTAA CGGACAACTG CATTGCTTG TAGAGCCACA 600  
 30 AGAAGCAACT TTAAATAAGG TTTACGGTTG CATTTTGATA CAACAACCGA TTAATAAGTC 660  
 ATGCTTTCCA CTTTGCGGGT TAGCATGACT TACCTAATAG ATAGAGCTAT TAGGTTTCAGC 720  
 TTCTAAAAAA TTACAGTTTT AGAGGAATAC AGTTGcTTGc tTCGCAACAA CTGCATAAGA 780  
 35 GCCATGGTTT TCGCTTTTGC GAATTAGCAT GACTTACCTA CTAGATAGAG CTATTAGGTT 840  
 CATCTTCTAA AAAATTACAG GTTTAGAGGA ATACAGTTGT TTGcTTGCA ACAACTGCAT 900  
 40 AAGAGCCTCT AGTAATTAAA ATTACAGAGG CTCTAAAAAT ACATCTAAAG GAGTGTCTGTA 960  
 TGAATCGGCA GGTATAGAA TTTTCTAAGT ATAATCCTTC GGGGAATATG ACGATACTTG 1020  
 TTCAITCAAA ACATGATGCT AGTGAATATG CATCTATCGC CAATCAGTTG ATGGCCGCAA 1080  
 45 CACATGTATG CTGTGAACAG GTAGGCTTTA TAGrATCAAC ACAAATGAT GATGGTAATG 1140  
 ATTTTCACTT AGTTATGAGC GGTAATGAAT TTTGCGGTAA TCGGACGATG TCATATATAC 1200  
 ATCATTTGCA GGAAAGTCAT TTGCTTAAAG ACCAACAGTT TAAGGTGAAG GTGTCTGGCT 1260  
 50 GTTCGGATTT AGTGCAATGC GCAATTCATG ATTGCCAATA CTATGAAGTT CAAATGCCAC 1320  
 AAGCCCATCG TGTGTGCCA ACAACAATTA ATATGGGTAA TCATTCATGG AAAGCAATAG 1380

55

	TTCAACATTT	GGTTGAAGCG	TTTGTGCGTG	AgcAACAATG	GAGTCACAAA	TATAAAACAG	1500
	TAGGTATGAT	GCTTTTGTGAT	GAACAACGTC	AATTTTTTACA	GCCATTAATC	TATATACCAG	1560
5	AAATTCAAAG	TTTAATTTGG	GAAAATAGCT	GTGGTTCTGG	TACAgCATCA	ATTGGGGTTT	1620
	TTAATAATTA	TCAACGTAAT	GACGCATGCA	AAGATTTTAC	AGTACATCAG	CCAGGGGGCA	1680
	GTATTTTAGT	GACATCAAAG	CGATGTCATC	AATTGGGATA	TCAAAC TTCA	ATTAAAGGAC	1740
10	AGGTTACAAC	TGTAGCTACA	GGaAAAAGCAT	ATATAGAATA	AGGAGCCTAC	AATGAATAAC	1800
	TTTAATAATG	AAATCAAATT	GATATTACAA	CAATATTTAG	AAAAGTTTGA	AGCGCATTAC	1860
	GAGCGTGTAT	TACAAGACGA	TCAATATATC	GAAGCATTAG	AAACATTGAT	GGATGACTAT	1920
15	AGTGAATTTA	TTTTAAATCC	TATTTATGAA	CAACAATTTA	ATGCTTGCGC	TGACGTTGAA	1980
	GAAAAAGCAC	AATTaATAAA	ATCACTGCAA	TATATTACAG	CGCAGTGTGT	TAAACAAGTG	2040
20	GAAGTCATTA	GAGCGAGACG	TCTATTAGAC	GGACAGGCGT	CTACCACAGG	TTACTTTGAC	2100
	AATATAGAAC	ATTGTATTGA	TGAAGAGTTT	GGACAATGTA	GTATAGCTAG	CAATGACAAA	2160
	TTATTGTTAG	TTGGTTCAGG	TGCATATCCA	ATGACGTTAA	TTCAAGTAGC	AAAAGAAACA	2220
25	GGTGCTTCAG	TTATCGGTAT	TGATATTGAT	CCACAAGCCG	TTGACCTAGG	GCGCAGAATC	2280
	GTTAACGTCT	TAGCACCAAA	TGAAGATATA	ACAATTACGG	ATCAAAAGGT	ATCTGAAC TT	2340
	AAAGATATCA	AAGATGTGAC	GCATATCATA	TTCAGCTCGA	CAATTCCTTT	AAAGTACAGC	2400
30	ATTTTAGAAG	AATTATATGA	TTTAACAAAT	GAAAATGTCG	TAGTTGCAAT	GCGCTTTGGT	2460
	GATGGCATCA	AAGCAATATT	TAATTATCCG	TCACAAGAAA	CAGCGGAAGA	TAAGTGGCAA	2520
	TGTGTGAATA	AACATATGAG	ACCACAGCAA	ATTTTGTGATA	TAGCACTTTA	TAAAAAAGCA	2580
35	GCTATAAAGG	TAGGTATTAC	GGATGTCTAA	ATTATTAATG	ATAGGCACTG	GTCCgGTCCG	2640
	AATGCAATTA	GCGAATATTT	GCTATTTAAA	ATCAGATTAT	GAGATTGATA	TGGTTGGACG	2700
40	TGCCTCAACA	TCAGAAAAAT	CAAAACGCTT	ATATCAAGCG	TATAAAAAAG	AGAAACAATT	2760
	TGAAGTCAAA	ATACAAAACG	AGGCGCATCA	ACATCTGGAA	GGTAAGTTTG	AAATTAATCG	2820
	TTTGATATAA	GATGTTAAAA	ACGTTAAGGG	TGAATACGAA	ACGGTTGTCA	TGGCATGCAC	2880
45	AGCAGATGCT	TATTATGACA	CACTACAGCA	ATTGTGCTTA	GAAACTTTGC	AAAGTGTCAA	2940
	ACATGTCATT	TTAATATCAC	CGACATTTGG	TTCGCAAATG	ATTGTCGAAC	AATTTATGTC	3000
	TAAATTTAAT	AAAGATATCG	AAGTGATTTC	ATTCTCAACT	TATCTTGCGC	ATACACGTAT	3060
50	TGTTGATAAA	GAAGCGCCTA	ATCATGTGTT	GACAACAGGT	GTAAAAAAGA	AATTGTACAT	3120
	GGGATCGACA	CATTCAAAC T	CAACAATGTG	TCAACGAATC	TCTGCTTTAG	CTGAGCAATT	3180

55

	TTATGTGCAC	CCACCACTAT	TTATGAATGA	CTTTTCATTG	AAAGCCATTT	TCGAAGGAAC	3300
	AGATGTACCG	GTTTATGTGT	ATAAGTTATT	TCCTGAAGGA	CCGATAACGA	TGACACTAAT	3360
5	CCGTGAAATG	CGTTTAAATGT	GGAAGGAAAT	GATGGTTATT	TTACAAGCAT	TTAGAGTGCC	3420
	GTCAGTCAAC	CTGCTTCAAT	TTATGGTGAA	GGAAAAATTAT	CCAGTACGTC	CTGAAACTTT	3480
	GGATGAAGGT	GATATTGAGC	ATTTTCGAAAT	CTTGCCAGAT	ATCTTACAAG	AATATCTGCT	3540
10	TTATGTAAGA	TATACCGCAA	TCCTCATTGA	TCCATTTTCA	CAGCCAGACG	AAAACGGACA	3600
	TTACTTTGAT	TTTTCAGCTG	TACCATTAA	GCAAGTCTAT	AAAAATGAAC	AGGATGTTGT	3660
15	TCAAATTCCA	AGAATGCCAA	GTGAAGATTA	TTACAGAACG	GCGATGATTC	AGCATATTGG	3720
	GAAAATGCTA	GGTATCAAAA	CGCCAATGAT	TGATCAGTTC	CTAACTCGCT	ATGAAGCAAG	3780
	TTGCCAGGCG	TACAAGGATA	TGCATCAAGA	TCAACACTTA	TCTTCTCAAT	TTAATACAAA	3840
20	TCTATTTGAA	GGAGATAAAG	CACTCGTCAC	AAAATTTTTG	GAAATCAATA	GAACGCTTTC	3900
	ATAATAAGGG	TTTGAAGTTT	TATAATAGAA	AAAAATTATT	GAATTATGTT	TGACATTTAC	3960
	ATAAAAATAA	GCAAATAATT	GAGAAAAATA	ATCATTACGA	TTTGATTAAG	TAATGCAACT	4020
25	TATCAATTTA	GAAAGAGGAA	AAGCAAATGA	GAAACTAAC	TAAATGAGT	GCAATGTTAC	4080
	TTGCATCAGG	GCTAATTTTA	ACTGGTTGTG	GCGGTAATAA	AGGTTTAGAG	GAGAAAAAAG	4140
	AAAACAAGCA	ATTAACGTAT	ACGACGGTTA	AAGATATCGG	TGATATGAAT	CCGCATGTTT	4200
30	ACGGTGGATC	AATGTCTGCT	GAAAGTATGA	TATACGAGCC	GCTTGTACGT	AACACGAAAG	4260
	ATGGTATTAA	GCCTTTACTA	GCTAAAAAGT	GGGATGTGTC	TGAAGATGGG	AAGACATACA	4320
35	CGTTCCATTT	GAGAGATGAC	GTAAATTCC	ATGATGGTAC	GCCATTTGca	TGctGACGCA	4380
	GTAAAGAAAA	ATATTGACGC	AgTTCAAGAA	AACAAAAAAT	TGCATTCTTG	GTAAAGATT	4440
	TCGACATTAA	TTGACAATGT	TAAAGTTAAA	GATAAGTACA	CGGTTGAATT	GAATTTGAAA	4500
40	GAAGCATATC	AACCTGCATT	GGCTGAATTA	GCGATGCCTC	GTCCATATGT	ATTTGTGTCT	4560
	CCAAAAGACT	TtAAAAACGG	TACAACAAAA	GATGGCGTTA	AAAAGTTCGA	TGGTACTGGT	4620
	CCATTTAAAT	TAGGTGAACA	CAAAAAAGAT	GAGTCTGCAG	ACTTTAACAA	AAATGATCAA	4680
45	TACTGGGGCG	AAAAGTCTAA	ACTTAACAAA	GTACAAGCAA	AAGTAATGCC	TGCTGGTGAA	4740
	ACAGCATTCC	TATCAATGAA	AAAAGGTGAA	ACGAACTTTG	CCTTCACAGA	TGATAGAGGT	4800
	ACAGATAGCT	TAGACAAAGA	CTCTTTAAAA	CAATTGAAAG	ATACAGGTGA	CTATCAAGTT	4860
50	AAGCGTAGTC	AACCTATGAA	TACGAAAATG	TTAGTTGTCA	ATTCTGGTAA	AAAAGATAAC	4920
	GCTGTGAGTG	ACAAAACAGT	CAGACAAGCG	ATTGGTCATA	TGGTAAACAG	AGATAAAATT	4980
55							

EP 0 786 519 A2

	ACAGACATTA ATTTTCGATAT GCCAACACGT AAGTATGACC TTAAAAAAGC AGAATCATTA	5100
	TTAGATGAAG CTGGTTGGAA GAAAGGTAAA GACAGCGATG TTCGTCAAAA AGATGGTAAA	5160
5	AACCTTGAAA TGGCAATGTA CTATGACAAA GGTTCCTTCAA GTCAAAAAGA ACAAGCAGAA	5220
	TACTTACAAG CAGAATTTAA GAAAATGGGT ATTAAGTTAA ACATCAATGG CGAAACATCA	5280
	GATAAAATTG CTGAACGTCG TACTTCTGGT GATTATGACT TAATGTTCAA CCAAACCTGG	5340
10	GGATTATTGT ACGATCCACA AAGTACTATT GCAGCATTTA AAGAGAAAAA TGGTTATGAA	5400
	AGTGCAACAT CAGGCATTGA GAACAAAGAT AAAATATACA ACAGCATTGA TGACGCATTT	5460
	AAAATCCAAA ACGGTAAAGA GCGTTCAGAC GCTTATAAAA ACATTTTGAA ACAAATTGAT	5520
15	GATGAAGGTA TCTTTATCCC TATTTACAC GGTAGTATGA CAGTTGTTGC ACCaAAAGAT	5580
	TTAGAAAAAG TATCATTAC ACAATCACAG TATGAATTAC CATTCAATGA AATGCAGTAT	5640
20	AAATAAAGGA GCAATTAGAT GTTCAAATTT ATCTTAAAC GTATTGCGCT CATGTTTCCA	5700
	TTGATGATTG TAGTAAGTTT TATGACATTT CTATTGACGT ATATTACAAA TGAAAATCCA	5760
	GCTGTGACAA TTTTACATGC ACAAGGGACG CCAAATGTAA CACCAGAGTT GATTGCAGAA	5820
25	ACGAATGAGA AGTACGGTTT CAATGATCCA TTATTAATTC AATATAAAAA TTGGTTACTT	5880
	GAAGCGATGC AATTTAATTT TGGTACAAGC TACATTACAG GTGACCCAGT TGCTGAACGT	5940
	ATTGGTCCAG CATTTATGAA TACATTGAAA TTAACAATAA TTTCAAGTGT TATGGTGATG	6000
30	ATTACATCAA TTATTTTAGG TGTAGTTAGT GCATTAAAAA GAGGAAAGTT CACTGATCGT	6060
	GCGATACGTT CAGTGGCTTT CTTTCTAACT GCATTACCAT CATATTGGAT AGCTTCAATA	6120
	CTTATTATTT ACGTTTCAGT GAAGTTAAAC ATATTGCCGA CTTCTGGATT AACAGGTCCA	6180
35	GAAAGTTACA TATTGCCAGT GATCGTTATT ACGATTGCCT ATGCTGGTAT TTACTTTAGA	6240
	AATGTTAGAC GCTCGATGGT GGAACAATTA AATGAAGATT ATGTACTTTA TTAAAGAGCA	6300
40	AGCGGTGTGA AATCTATCAC ATTAATGTTG CATGTGTTGC GTAATGCTTT ACAAGTTGCG	6360
	GTATCAATCT TTTGTATGTC TATACCAATG ATAATGGGTG GACTAGTTGT TATCGAGTAT	6420
	ATCTTTGCAT GGCCTGGACT AGGTCAATTA AGTTTAAAAG CAATACTTGA ACACGATTTT	6480
45	CCAGTCATTC AAGCATATGT ATTAATTGTA GCGGTATTAT TTATTGTATT TAATACATTA	6540
	GCAGATATCA TTAATGCGCT ATTAAATCCA AGATTAAAGG aGGGCGCACG ATGATAATTT	6600
	TAAAmCGATT ATImCArGwT AAAGGTGCAG TAATTGCTTT AGGCATTATT GTATTATATG	6660
50	TCTTTTAGG ATTAGCAGCA CCACTTGTGA CATTTTATGA TCCTAACCAT ATCGATACAG	6720
	CAACAAATT TGCTGGCATG AGTTTTCAAC ATCTACTAGG TACTGACCAT TTAGGTAGAG	6780

55

	TATTTGTTTC TGTACTTATT GGATCTATTT TAGGATTCTT ATCAGGATAT TTCCAAGGGT	6900
	TTGTTGACGC CTTAATCATG CGTGCGTGTG ATGTTATGTT GGCATTCCCA AGTTATGTTG	6960
5	TAACGTTAGC ATTAATTGCA TTGTTTGGAA TGGGTGCCGA AAATATTATC ATGGCATTTA	7020
	TTTTGACGCG TTGGGCATGG TTCTGTCGTG TTATACGTAC AAGTGTTATG CAGTACACTG	7080
	CTTCTGACCA TGTAAGATTT GCTAAAACAA TCGGTATGAA TGATATGAAA ATTATTACACA	7140
10	AACATATTAT GCCATTAACA TTAGCAGATA TTGCTATCAT CTCTAGTAGC TCGATGTGTT	7200
	CAATGATCTT GCAAATATCT GGCTTTTCAT TTTTAGGATT AGGTGTCAA A GCGCCTACTG	7260
	CAGAGTGGGG CATGATGCTT AACGAaGCTA GAAAAGTGAT GTTTACACAT CCTGAAATGA	7320
15	TGTTTGCGCC AGGTATTGCC ATAGTGATTA TAGTGATGGC ATTTAACTTC TTATCCGATG	7380
	CTTTACAAAT TGCTATTGAT CCCC GCATCT CTCTAAAGA TAACTTCGT TCTGTGAAAA	7440
	AAGGAGTGGT GCAATCATGA CATTGTTAAC AGTTAAACAT TTGACGATTA CAGATACCTG	7500
20	GACAGATCAA CCACTCGTGA GTGATGTGAA TTTTACATTA ACTAAGGGTG AAaCTTTAGG	7560
	CGTTATTGGA GAAAGTGGA GTGGTAAATC AATCACTTGT AAATCGATTA TTGGTTTGAA	7620
25	TCCCGAACGA CTCGGGGTGA CAGGTGAAAT TATCTTTGAT GGTACatCAA TGTGTGCATT	7680
	ATCTGAATCG CAATTGAAAA AGTACCGTGG TAAAGACATT GCGATGGTCA TGCAACAAGG	7740
	TAGTCGTGCC TTTGACCCAT CAACTACTGT CGGTAAACAA ATGTTTGAGA CTATGAAAGT	7800
30	ACATACGTCA ATGTCTACAC AAGAAATTGA AAAGACATTG ATTGAATATA TGGATTATTT	7860
	AAGTTTGAAA GATCCTAAAC GTATATTAAA ATCATACCCT TACATGTTAT CAGGAGGAAT	7920
	GTTACAGCGA TTGATGATTG CTTTAGCGTT AgcTTTgAAA CCAAAGTTAA TCATTGCTGA	7980
35	TGAGCCGACA ACGGCTTTAG ATACAATTAC ACAATATGAT GTA CTGGAAG CATTTATAGA	8040
	TATTTAAAAA CACTTTGACT GTGCGATGAT TTTCA TTTCa CATGATTTAA CGGTTATTAA	8100
	CAAGATTGCA GACCGTGTTG TTGTGATGAA AAATGGTCAG CTTATTGAAC AAGGGACACG	8160
40	TGAATCAGTC TTGCATCATC CAGAACATGT TTATACGAtt ATTKTATTAT CAACGAAGAA	8220
	GAAGATTAAT GATCATTTTA AACATGTGAT GAGGGGTGAT GTACATGATT AAAATTAAAG	8280
45	ATGTTGAAAA GTCATATCAA AGCGCACATG TTTTAAAGCG TCGTCGAACA CCTATCGTGA	8340
	AAGGTGTGTC ATTTGAGTGT CCAATCGGTG CGACGATTGC GATTATCGGA GAAAGTGGA	8400
	GCGGTAAATC GACGTTGAGT CkAtATGATAT TAGGTATTGA GAAACCGGAT AAAGGTTGTG	8460
50	TAACCTTAAA TGATCAACCG ATGCATAAGA AGAAAGTGAG ACGTCATCAA ATTGGTGCTG	8520
	TATTTCAAGA TTATACGTCA TCATTACATC CATTTCAGAC TGTTAGAGAA ATCTTATTTG	8580
55		

EP 0 786 519 A2

	TGTTGGAAGA AGTCGGTCTA TCTAAGGCAT ACATGGATAA ATATCCTAAT ATGTTATCAG	8700
	GTGGAGAGGC GCAACGTGTT GCGATTGCGC GTGCAATATG TATTAACCCT AAATATATTT	8760
5	TGTTTGATGA AGCCATTAGT TCACTCGACA TGTC AATTCA AACACAAATA TTAGATTTAT	8820
	TGATTCATTT ACGTGAAACG CGTCAGTTGA GTTATATTTT TATCACACAT GATATTCAAG	8880
	CTGCCACGTA TTTATGTGAT CAATTAATTA TTTTAAAAA CGGAAAAATA GAAGAACAAA	8940
10	TTCCGACAAG CGCATTGCAT AAAAGTGACA ATGCTTATAC AAGAGAATTA ATAGAAAAAC	9000
	AACATCAATT CTAAGGAGTG AGATAATGAA AGGTGCAATG GCTTGCCCT TTTTGAGATT	9060
	ATATATATTA ACATTGATGT TCTTTAGTGC CAATGCAATC TTAAACGTGT TTATACCTTT	9120
15	ACGAGGGCAT GATTTAGGCG CAACGAATAC GGTATCGGT ATCGTTATGG GGGCATACAT	9180
	GTTAACAGCA ATGGTATTTC GACCATGGGC AGGACAAATT ATTGCTCGTG TCGGTCCCAT	9240
20	TAAAGTATTA AGAATTATTT TGATTATCAA TGCCATAGCT TTAATTATTT ATGGTTTTAC	9300
	TGGCTTAGAA GGTATTTTCG TAGCACGTGT TATGCAAGGT GTGTGTACGG CATTCTTTTC	9360
	TATGTCTTTA CAGCTAGGTA TTATTGATGC ATTACCAGAG GAACATCGTT CTGAAGGTGT	9420
25	ATCATTGTAC TCGCTATTTT CAACGATTCC AAACCTAATC GGACCATTAG TTGCCGTAGG	9480
	TATTTGGAAT GCAAATAATA TTTCATTATT TGCAATTGTC ATTATCTTTA TCGCATTAAAC	9540
	AACAACATTC TTTGSTATCG CGTGACCTTT GCTGAACAGG AACCCGATAC GTCAGATAAG	9600
30	ATTGAAAAAA TGCCGTTTAA CGCTGTAAC TTTTTCGCGC AATTTTTCAA AAATAAAGAG	9660
	TTGTTGAACA GTGGTATTAT CATGATTGTT GCATCGATTG TATTTGGTGC AGTTAGTACA	9720
	TTGTACCGT TATACACAGT GAGTTTAGGA TTCGCGAATG CGGGAATCTT TTTGACAATA	9780
35	CAGGCCATCG CAGTTGTTGC GGCAAGATTT TACTTAAGGA AATACATTCC GTCAGATGGT	9840
	ATGTGGCATC CTAAATATAT GGTATCTGTA CTATCATTAT TAGTAATCGC GTCATTTGTA	9900
	GTGGCATTG GTCCGCAAGT AGGTGCAATT ATTTTCTATG GTAGTGCAT ATTAATAGGA	9960
40	ATGACGCAAG CAATGGTGTA CCCAACATTA ACATCATACT TAAGCTTCGT CTTACCAAAA	10020
	GTAGGTCGTA ATATGTTGTT AGGTTTATTT ATTGCCTGTG CAGACTTAGG TATATCGTTA	10080
	GGTGGCGCAT TGATGGGACC TATTTCCGAT TTAGTAGGAT TTAAATGGAT GTATCTAATT	10140
45	TGTGGTATGT TAGTCATTGT AATAATGATT ATGAGTTTCT TGAAAAAGCC AACACCACGT	10200
	CCAGCGAGTA GTCTTTAATG AAGTGAATTA AAGCATATTA AGTTAATGAA TATTTAAATT	10260
50	TTAAAAGGTA TATTGAGCAT GGCGATTCAT GTGCTTCATG CTAGGACATG AAACATTCTA	10320
	TATGGCTCGT TTTTAGAACG ACATATATCT AAATAAGCA CGCTTAAG TGAGTTTTGA	10380

55



EP 0 786 519 A2

	TTACATGAAA ATATGCAAAA CGAGTATAAC TGCTAATTGA TAGAAATAGC TCACCATAAA	10500
	ATTACGGTAT GATTTTAAAT ATAAGTAAGT CGCACTACCT GCTAGTATCA ATGCTGGAAT	10560
5	GAATTCACAC CATGTATTAA TGTATGGATA GTAGAACAGA GTTTCAGGA TAATGGACAA	10620
	TACTATTGTA ATCTTTAAAG GTATTAATCT GCTTAATTCT TGAATTAAAA TATGACGGAA	10680
	AATAAGTTGA CAAATCAAAG TATTTAATAT AATGGTTAAC GAAAATATAG CTATTAAACT	10740
10	GATGGAaCCA TACCCTTTAA TGAGCGGGTA AATGTCAAAG ACAGTAAAGG AATCTACATT	10800
	TAGTGCgAAA ATATTGAAAT GATTTAAAAG TAAAAAGAGT ACGACACTTA GTGTAAATGA	10860
15	TATAAGAATA TGCCATTTAT ATTTAGCACT AGCAACGATT TGCGAACGTA TCATTGGAAT	10920
	AAACGCATCT TCATGCATCA GACGAAAAAT AGCTAGTGAA ATAATAACTG CGAGTAAATA	10980
	GCTAATGTTT ATTGAAATAG GAAAAGAGAA ACCCCACGGA GCTTGTGAG TGAATACAGC	11040
20	TACTAACCCA AAAGTTAAAA AGACGATAAT GATCGGCAAG ATGTTAACCA AAAATATGTA	11100
	AAGGAAAAATA AATCCAATAT CACGTTTGAA AAAACGCGAT TGTTCCGGTAG CGTATTCTTC	11160
	TTCTATGTAA TGTTTATTG TATTTGACAT AGTATACCTC TTAAATAGTT GTATTATATA	11220
25	GATACTTTAG CACATATTAC TTTGTATTGT ATGTTTTATA CATTAAAATT TAAAATGAAA	11280
	AACATATCAT AAAATGTTT TATAAAATGA AGCGCTTCCA TTGTGTTTTG TTTTGTAAGG	11340
	TGTATCATAA ATATTGAATT GAAATTTTGG GGGGAGGTAT TGTAATGACG TTTCTTACAG	11400
30	TCATGCAATT TATAGTTAAC ATTATCGTTG TAGGATTCAT GCTTACGGTT ATTGTTATCG	11460
	GGCTTATTG GTTAATTAAA GATAAAAGAC AATCACAACA TAGTGTATTA AGGAATTATC	11520
	CTTTACTAGC ACGTATTAGA TATATTTTCCG AAAAAATGGG ACCGGAATTA CGTCAGTATT	11580
35	TATTTTCTGG GGATAATGAA GGGAAACCTT TTTACGTAA TGATTATAAA AATATCGTTT	11640
	TGGCTGGAAA ATATAACTCT CGTATGACCA GCTTCGGTAC TACTAAAGAT TATCAAGACG	11700
40	GCTTTTACAT ACAGAACACA ATGTTTCCGA TGCAACGTAA TGAGATTTCG GTAGATAATA	11760
	CAACATTGTT ATCAACATTC ATTTATAAAA TCGCGAATGA GCGTTTATTT AGTCGTGAAG	11820
	AATATCGTGT GCCGACAAAG ATTGATCCGT ATTACTTAAG TGATGACCAT GCAATAAAAT	11880
45	TAGGTGAACA TTTAAAACAT CCATTTATTT TAAAACGTAT CGTAGGACAA TCTGGTATGA	11940
	GTTATGGCGC TTTAGGAAAA AATGCCATTA CAGCTTTATC TAAAGGTCTA GCTAAAGCGG	12000
	GCACTTGGAT GAATACAGGT GAAGGTGGCT TATCAGAATA TCATTTAAAA GGTAATGGGG	12060
50	ATATCATTTT CCAAATTGGT CCCGGTTTAT TTGGTGTTTC TGATAAAGAA GGTAATTTTA	12120
	GTGAAGGTTT ATTTAAAGAG GTTGCACAGT TATCTAACGT ACGCGCATTT GAGCTGAAGT	12180

55

	TTGCTAAAAT	CCGAAATGTT	GAACCTTATA	AAACAATCAA	TTCACCTAAC	CGTTACGAAT	12300
	TTATTCATAA	TGCTGAAGAT	TTGATTCGTT	TCGTCGATCA	GTTGCAGCAA	TTAGGTCAAA	12360
5	AACCACTAGG	ATTCAAAATT	GTAGTAAGCA	AAGTTTCAGA	AATTGAAACA	CTTGTACGTA	12420
	CGATGGTGGA	ACTAGATAAG	TATCCAAGCT	TTATTACGAT	TGATGGTGGT	GAAGGTGGTA	12480
	CTGGTGCAAC	ATTCCAAGAA	TTACAAGATG	GTGTTGGCTT	ACCGCTATTT	ACAGCTCTAC	12540
10	CTATTGTGTC	TGGCATGTTA	GAAAAATATG	GTATTTCGAGA	TAAAGTGAAA	TTGGCGGCAT	12600
	CTGGTAAGTT	AGTGACACCA	GATAAAATTG	CGATTGCACT	AGGTTTAGGT	GCAGATTTTG	12660
15	TAAATATCGC	ACGTGGGATG	ATGATTAGTG	TCGGTTGTAT	AATGAGTCAA	CAATGTCACA	12720
	TGAATACGTG	TCCTGTAGGT	GTTGCAACGA	CAGATGCGAA	GAAAGAAAAA	GCATTGATTG	12780
	TTGGAGAAAA	GCAATATCGT	GTCACAACT	ATGTAACAAG	TTTGCATGAA	GGCTTATTCA	12840
20	ATATTGCAGC	AGCTGTTGGC	GTATCCAGTC	CTACAGAAAT	TACTGCTGAT	CATATTGTAT	12900
	ATCGAAAAGT	CGATGGTGAG	TTACAAACGA	TACATGATTA	TAAATTAAAA	CTCATTAGTT	12960
	AACTTAATTA	TTTCGGGAAA	TTGAAAGCAG	CGGATTTTAG	CGTTACTGCA	AATAATTTTA	13020
25	TATTAGTAGT	GGATGCTGGT	CACACAAGAA	CTTCAAATAT	TAAAGCCCTC	AGAATATGAA	13080
	TTAAGGTTTG	TAACCTTAGT	CTTATCTGAG	GGCATTTTTA	AGTTATAAAC	TATTTGTCGT	13140
	CCATTTTATC	TTTTCTTTT	AAACCTCTGT	GCTTTAATTG	CTTTTCAAGT	TTTTCAAAAC	13200
30	TAATATCTTT	ATTTTCTTTA	GTCGAAACAC	CAAGACGTTT	ATTTAATTTT	TTCATGTCAA	13260
	CTTCTGTGTA	ATCTATGTCT	AAGTGyTCAA	TTGCTTTTTT	ATCTTTATAG	TCTACTTTGT	13320
	ATTTTACGCC	TTAAGGTCT	TTGAAAATAC	TTTCAGATTT	GGCGAATAAC	TTTTTGGCTT	13380
35	CGTCTTTATC	CATACCTAGA	TCGTCATATT	TAATTGTGTT	GATTGTAGAC	TGTTTTAAAA	13440
	CTTTATCATC	TTTATATGTG	ATAGAAGTTA	GTACATGTTT	ACCACTAACA	TCACCWTCAT	13500
40	ATGTTTTGGT	TTGTTCTTTA	CCACAAGCTG	ATAATGCAAT	GATACAACT	AATGCTACTA	13560
	CAATTAATGA	ACATAATTTT	TTCAAAGTCA	GTCGCCTTCT	TTGATATTTT	GTATTATAAA	13620
	GAAATTATAA	CATTTACTAA	AAAATGATGT	TATTCAAAAA	TTTAAATTTT	GTCATTTTTT	13680
45	TTGAAGATAT	GAGTTTTTTT	AAGCGGATTC	CTCACAAAAT	TTTAAAAATA	TTTAAGCCTk	13740
	AAAATGATAA	AGCGkTAGGG	AACGTTTTTC	TGAAAGTTAG	TGATACAATA	GTTTTAAGTT	13800
	GAAATACAGG	AGGATGAATA	ACATGAATCA	GTCAGTCAAA	TTACTTAAAC	ATTTAACAGA	13860
50	TGTAAACGGC	ATTGCTGGTT	ATGAAATGCA	AGTTAAAGAA	GCAATGCGTa	ACTATATAGA	13920
	GCCTGTCAGT	GATCAAATTA	TTGAAGATAA	CTTGGGTGGC	ATTTTTGGAA	AGAAAAATGC	13980
55							

	AACAAAGATT GATAAACATG GTTTTATTTT ATTTACGCCA kTgGTGGATG GTGGAATCAA	14100
	GTCATGCTAT CTCAAAAAGT AACGATTACA ACAGATTCCG GCAAAGAAAT TAGAGGTATC	14160
5	ATCGGTCTTA AACCGCCACA TGTCTTAACG CCTGAAGAAC GTAAAAAGCC AATGGAAATC	14220
	AAAAATATGT TTATAGATAT TGGTGTTAGT AGCAAGGAAG AAGCTGAAGA AGCTGGCGTT	14280
	GAAGTAGGCA ATATGGTTAC GCCATATAGT GAATTTGAAG TGCTTGCAA TGATAAATAT	14340
10	TTAACTGCGA A <sub>r</sub> CATTTGAT AATCGCTATG GCTGTGCATT AGCTATTGAG GTATTAAAC	14400
	GTTTAAAAGA TGAAAATATT GGCATTA <sub>A</sub> CT TATACAGTGG TGCCACAGTG CAAGAAGAAG	14460
15	TTGGTTTGCG TGGTGCGAAA GTGGCAGCGA ATACGATTAA ACCAGACTTG GCGATAgcTG	14520
	TcGATGTAGG TATTGCTTAT GATACCCAG GTATGTCAGG TCAAACGAGC GATAGTAAAC	14580
	TAGGCGGTGG TCCAGTTGTC ATTATGATGG ATGCTACAAG TATTGCTCAC CAAGGTTTGC	14640
20	GAAAgcATaT TAAAGATGTA GCTAAGGAAC ATAACATCGA AGTACAATGG GATACGACAC	14700
	CAGGTGGAGG TACAGATGCG GGAAGTATTC ATGTCGCAA TGAAGGTATT CCAACGATGA	14760
	CAATCGGTGT TACGCTGCGA TACATGCATT CTAATGTTTC AGTGCTCAAT GTAGATGATT	14820
25	ATGAAAATTC TATCCGTCTT GTTACTGAAA TTGTCCGTTT ATTGAATGAT GAAAGTTATA	14880
	AAAATATCAT GTGGTAATCA AATCCATAAA TAATAAGAA TCCTTTTAAT ATGGTAGGTT	14940
	GTTAAACAAT TGTCTAATTT TAATTCCTAG TCATTAGACA GTATCCATGT TAATAGGATT	15000
30	TTTTGTTTTT AATT <sub>T</sub> AAATG CTGAAAATCA ATTATGCCTA AATTTTGATA TTACAAGAAA	15060
	ATGATTTTTT CTAAATGTA ATTGCACTAA AAACCAAAAA AACGGGAATA ATATACCTGA	15120
	TATATTACAT GAGGAGCGGT GCAAATGTTG TTAGAAATTA AAGATTTAGT GTATAAAGCG	15180
35	AGCGATAGAA TCATACTAGA TCATATCAGT CTAAAAGTAG ATAAAGGCGA GAGTATTGCC	15240
	ATTATAGGTC CATCAGGTAG TGGTAAAAGT ACATTTCAAA AGCAAATATG TAATTTGTTT	15300
40	AGTCCA <sub>A</sub> CTA GTGGAGA <sub>A</sub> CT TTATTTTAAA GGTAAACCTT ATAATGATTA TGACCCGGAA	15360
	GAATTGCGTC AACGAATCAG TTATTTGATG CAGCAAAGTG ACTTGTTTGG TGAAACGATT	15420
	GAAGATAACA TGATATTCCC ATCACTTGCA CGTAATGATA AATTTGATAG AAAACGTGCA	15480
45	AAGCAATTAA TTAAAGATGT CGGTTTG <sub>G</sub> GA CATTATCAAT TAAGTTCCGA AGTGGAAAAT	15540
	ATGTCGGGTG GTGAGCGGCA AAGAATTGCT ATAGCGCGCC AACTGATGTA TACACCGGAT	15600
	ATTCTTTTAT TAGATGAATC GACCAGTGCA TTAGACGTTA ATAATAAGA AAAGATAGAA	15660
50	AATATCATTT TTAAATTAGC AGATCAAGGC GTGGCAATTA TGTGGATTAC CCACAGCGAT	15720
	GACCAAAGTA TGCGACACTT TCAAAGCGT ATAACAATTG TTGATGGTCA AATTTCTAAT	15780

55

EP 0 786 519 A2

	CATTCCGATT ATCATTTTCAT ATAAAGAAGG TTTACATATT ATTAAAGATT TAATTGTTGC	15900
	GACATTACGA GCAGTTGTGC AATTAATCAT TTTGGGATTI TTGCTGCATT ATATTTTAA	15960
5	AATAAACGAT AAATGGCTGC TTATTTTATG TGTATTGGTC ATTATTATTA ATGCATCATG	16020
	GAATACAATT AGTCGAGCAT CACCAGTGAT GCATCATGTG TTTTGGATAT CATTCTAGC	16080
	TATCTTCATT GGAACGGCAT TACCGCTTGC AGGTACTATT GCGACAGGGG CCATTCAATT	16140
10	TACCGCAAAT GAAGTTATAC CTATCGGCGG CATGCTTGCA AATAATGGCT TGATTGCAAT	16200
	TAATTTAGCT TACCAGAATT TAGATCGTGC ATTCGTACAA GATGGTACTA ATATTGAATC	16260
	TAAATTATCA CTTGCAGCTA CACCTAAATT GGCTTCTAAA GGTGCAATAC GTGAAAGTAT	16320
15	TCGTTTAGCT ATAGTGCCAA CTATTGATTC GGTAAAACA TATGGGCTTG TGTCGATTCC	16380
	TGGTATGATG ACAGGCTTAA TTATTGGTGG CGTACCACCT TTACAAGCGA TTAAATTTCA	16440
20	ATTGTTAGTC GTGTTTATTC ATACAACGTC GACCATTATG TCTGCTTTGA TTGCGACATA	16500
	TTTAAGCTAT GGTCAATTTT TCAATGCAAG ACATCAATTA GTAGCACGAA ATACTGATGT	16560
	TAAGAGTGAA TCATGATAGA TTTTACTGCA TCAGATTTAG GCATTAGTTT TAATTGAAA	16620
25	TGAAGTGACG CGCACATATA GTATCGCTAT TCATTAGCGC AGCGAAAATA TTCATAAAGG	16680
	CACGCATACT TTGTAGTCAG TTATCTGTTC TGACATATAA AGCGTGCGTG CTTTTTTGGA	16740
	GTTATTGTTG AAAGTGAAGT AATTATACAT AATTATTAAA TGACATACTT GTGTTAATTT	16800
30	TTCAAATACT GAAAAACAAT TTCaATAATT TTCCaATTAA GCACAGAAAA TTAAAGCAAA	16860
	ATATTATATA ATAGAACGGT TATATATAaA nATTngTgCA CACATTTTTT AATAAATCGT	16920
	TATTCTAAGG GAAATGAATA TCGGAAATTT TGTTTGAAAG GAGTTTTTAAA TTGTCAATCA	16980
35	TGCGACTATT TACATTCATT TTAAGTATTT TTATCGTAGG AATGGTTGAA ATGATGGTTG	17040
	CAGGAATTAT GAACTTGATG AGTCAGGACT TACATGTATC AGAAGCTGTC GTTGGTCAAT	17100
	TAGTGACAAT GTACGCTTTA ACATTGCGA TATGTGGACC TATTCTGGTT AAATTAACGA	17160
40	ACCGTTTTTC ATCAAGGCCT GTATTATTAT GGACATTACT TATATTTATC ATTGGTAATG	17220
	GCATTATTGC TGTAGCGCCA AATTTTTCaA TATTAGTAGT TGGTAGAATT ATCTCATCTG	17280
45	CAGCAGCAGC ACTAATTATC GTAAAAGTAT TAGCTATTAC AGCGATGTTA TCAGCACCTA	17340
	AAAATCGTGG TAAAATGATT GGAATTGTCT ATACAGGGTT TAGTGGTGCT AATGTTTTTG	17400
	GTGTACCAAT TGGAACGGTT ATCGGCGATT TAGTAGGTTG GCGCTATACA TTTCTATTCT	17460
50	TAATTATTGT GAGTATTATT GTTGGCTTCT TGATGATGAT CTATTTACCG AAGGATCAGG	17520
	AAATACAACG AGGCCCTGTG AATCATGAGA CACCATCTCA TGAAAATCAT GTTACTTCGA	17580
55		

CAAACTCAGT GACATTCGTC TTTATAAATC CACTTATTTT ATCTAATGGT CATGATATGT 17700  
 CATTTCGTTTC ATTAGCACTT CTAGTAAATG GAATCGCTGG CGTTATTGGA ACATCATTAG 17760  
 5 GTGGTATATT CTCCGATAAA ATTACAAGTA AGCGTTGGTT AATGATTCTT GTTTCATATT 17820  
 TTATCGTCAT GATGTTACTT ATGAATTTAA TCTTACCTGG TTCAGGTCTA TTGTTAGCAG 17880  
 GACTATTTAT TTGGAATATC ATGCAATGGA GTACTAATCC AGCAGTGCAA AGCGGTGTGA 17940  
 10 TTCAACATGT TGAAGGCGAC ACAAGCCAAG TAATGAGTTG GAACATGTCT AGTTTAAACG 18000  
 CTGGTATTGG TGTGGAGGC ATTATTGGAG GCTTGGTCAT GACACATGTT TCTGTTCAAG 18060  
 CTATCACATA TACGAGTGCC ATCATTGGCG CATTAGGATT AATCGTTGTT TTCACATTGA 18120  
 AAAATAATCA TTATGCTAAA ACATTTAAAT CATCATAATT CTCATATGAm AAGCACGCCT 18180  
 GCTATCAAAT TCAGGTGTGC TTTTITAGAT GCGATAACGT TATTGATATG TGCATAATA 18240  
 20 GCGACGTTCA TTATGATACA TCGCCAAGG CATTTTACCG CTTTITAGCAA AATTAGCTAA 18300  
 ATCATTITGC ATTTGTCGAC TTAAAAATTT AAGGTGaGCA GTTGTGGaT ATgAT 18355

## (2) INFORMATION FOR SEQ ID NO: 68:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1192 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

CGCAAAGAAG TACAAAAAAT GTTTTACAA GAAGGTATTA AAACACCTCA ACCAATTATG 60  
 35 ACTGCTTATA ATCATAGTGA AAACGgTGTT TAGTAGTTTA TAATACATGG AGGTCATATT 120  
 TAATGGCGTC AAAATATGGA ATAAATGATA TAGTAGAAAT GAAAAACAA CATGCGTGTG 180  
 40 GAACAAACCG TTTTAAGATT ATTAGAATGG GTGCAGACAT AAGAATTAAA TGTGAAAATT 240  
 GTCAAAGAAG TATTATGATT CCACGTCAA CGTTTGATAA AAAACTTAAA AAAATCATCG 300  
 AATCTCATGA TGATACACAA AGATAGGAGA ATGATTAATG GCTTTAACAG CAGGTATCGT 360  
 45 TGGATTGCCA AACGTTGGTA AATCAACATT ATTTAATGCA ATAACAAAAG CAGGTGCTTT 420  
 AGCAGCGAAC TATCCATTGG CTACGATTGA TCCTAATGTA GGGATAGTAG AAGTGCCAGA 480  
 TGCTAGATTA CTTAAATTAG AAGAAATGGT TCAACCTAAA AAGACATTGC CGACTACATT 540  
 50 TGAATTTACA GATATCGCTG GTATTGTGAA AGGTGCTTCA AAGGGAGAAG GGTTAGGTAA 600  
 TAAATTCTTA TCACATATTA GAGAAGTAGA TGCGATTGTG CAGGTCGTTC GTGCATTGTA 660

55

TAATATGGAA TTAGTACTAG CGGACTTAGA ATCTGTTGAG AAACGTTTGC CTAGAATTGA 780  
 AAAATTAGCA CGTCAAAAAG ATAAGACTGC TGAAATGGAA GTACGTATTT TAACAACATAT 840  
 5 TAAAGAAGCT TTAGAAAATG GTAAACCCGC TCGTAGTATT GACTTTAATG AAGAAGATCA 900  
 AAAATGGGTG AATCAAGCGC AATTACTGAC TTCTAAAAA ATGCTTTATA TCGCTAATGT 960  
 TGGTGAAGAT GAAATTGGTG ATGATGATAA TGATAAAGTA AAAGCGATTG GTGAATATGC 1020  
 10 AGCGCAAGAA GACTCTGAAG TGATTGTTAT TAGTGCAAAA ATTGAAGAAG AAATTGCTAC 1080  
 ATTAGATGAT GAAGATAAAG AAATGTTCTT AGAAGaTTTA GGTATCGaAG AACCAGGATT 1140  
 15 AGATCgrTTA ATTAGGAmCA ctTATGAATT ATTAGGnTTA TCCACCATAA TT 1192

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7494 base pairs  
 (B) TYPE: nucleic acid  
 20 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

AATATAGCTG CAATAGCATC TCGTTTCATT TGTATAATCA ATTCCGGTTT AAATATCAGT 60  
 GTGAACGTAA GCACGACACA GATTAAAAAT AACACTGCCG GAATGAGTCG TTTCAATCGT 120  
 30 CGCTtCCAAA ACTCTAGCAA ATCGATTTTT TGCGTCCGAT AATACTCACT TATCAACAAA 180  
 CTTGTTATTA AATAACCTGA AATAACGAAG AATGTATCTA CTCCTAAAAA GCCCCCACTT 240  
 AACCATTGTG CATTCAAGTG ATAAATAATG ATTCCCTATAA CTGCGAATGC CCTCAATCCA 300  
 35 TCTAATCCAG GTAAGTATCG CGGGGAATAC ATTTTTTCTA AACGTTTAAA GTCTTTTGTA 360  
 TCCA<sup>1</sup>GTAA TAAACGCCCC ATTTATTTTT CTCTATTTTG TAGTATATCA CAATATTTTT 420  
 GAAA<sup>1</sup>A<sup>1</sup>AAAA TATTGCACTG aTTTTCA<sup>1</sup>TTA ATTGATTTAA CCCTTAATTA AGATAGTTTT 480  
 40 AAATTTTTTA TTAAGTAGAA AACAATTATT ACAGTTGATT TCATTACTGC AAACCACATA 540  
 TAAATTTGTC GATTTTACTA CATAACATAG ATTATCATAG ATTCTTGAAT TTTTAGCAAA 600  
 45 ATA<sup>1</sup>CTGTTA TTTTCATTAT ATTTTTACAA AAAAAGGTTT GTTTTATATT TTATGCATCT 660  
 TACTGTAACA GAATCATTAA GATATGCTAT TCGAATATAC TTTTCAAAA TTTATATAAT 720  
 GAATAAATTA ACATGTATTG AAAAAAAGC GAAATGCAGC CTATCCTCTA ATGTAAACCA 780  
 50 AACGATATAT CTCGTCAGAC TTTATATTTA AACGCTATGT GTCAC<sup>1</sup>TTT<sup>1</sup>TA AAATGAATAT 840  
 TACTAAGATT GTCATATCAA TTATTATTGC ATCGAATTAA TCTTTTAAAT TTCTGTAATA 900

55

	ACGGAAGTCA TTATTAGAAT AAAAATACTG TGCACTAATA AATTTATCAA TTGTTCTCTAA	1020
	ATAAATACCA TCGATATTTT GTTCTTTACA TGTCAATTATA ACTTTATCTA AAAGTTTTTTT	1080
5	ACCTATTTTT AAATTCCTAT AACCTTTATC AACAAACATT TTTTAAAGTG CAGACATATT	1140
	ATTATCTAGT CTAATCAAAC CTATAGTACC AACAATATTT TGaTGATTGT TTATTGCAAG	1200
	CCAAAATgCC CTCCATTATT CAAATAGTTA TGTTTCGATGT TCTCCAAATC AGGTTGATCA	1260
10	TCTCTATCAA TTTTATATa AATTCATTTT TTTGAATCGA TAAAATAAAC TCGATTAGCT	1320
	CTTCCTTATA AGACCTATTA TATTCAATTA TGTTTATAGC CATTTTATC TCCTTTTTCA	1380
	TTTAATTIAA TTATAAAATG TGCGTTTAGT TTGTATCTAG TGTACTCAGT ACAGCCTCAA	1440
15	ATGAAGTTTC ATTCCACTTG GCACTTAATA AAGACAAGTA TTTTAGCAGT AATACAATAA	1500
	AGTCCAATAA ATTTCCCTAA CTTCAATATC CACTTTTTAA AAAATGTATT TTTAATTAAT	1560
20	AAAAAACTC TCCCCAATTT CTATGGGAAG AGCTATATAT TTAATGTCTA AACATTACTT	1620
	TTATTTATTA TGAAGGAATT AGAATCCCCA AGCACCTAAA CCTTGTGCTT TGTATGCTTT	1680
	AACAGCTGCG TTGATTTGTT GGTCAACAGT GTTTGTTGGA CCCCAACCTG GCATAGTTTG	1740
25	GAATAAACCT GAAGCACCTG ATGGGTTGTA AGCATTTACT TGACCATTG ATTACAGAGC	1800
	GATGATTGCA GCCCATGTAG AAGCTGAAAC ACCAGTACGT TGAGCCATGA TTTGAGCTGC	1860
	TGATGAACCA GTAGCACCTG CAGTATTACC ATTGCTTAAT CTCACTGAAC TTGAAGTAGT	1920
30	TGAAGTGCTG TAGTTATGGT AAGTTGGAGC TGAACAGCT TCAACGTtTG AGTTACTTGA	1980
	TTGTGCATTG TAGCTTACTG ATTGTACATT TGAACCTTGG TTGTATGAAG TAGTGTAGTC	2040
	TGCACCTGCA ACGTTTGAGA AACCAGCAGT TTGACCATTA GCTGCTTCAT AGCTCCATGA	2100
35	CCATGTAGTA CCATTTGAAG TGAAGTTATA TTGGAACCA TCTTTTACAA AGTGGATGTC	2160
	ATAFGCACCA TCTTTGATTG GAGCTGCATT TAATTGATCT TGGTGATTAT GCGCTAAGTC	2220
	AACTAAGTGT GCTTGATCAA CGTTTACTTC AGCAGCGTGT GCTTGATGTC CTGTACCTGC	2280
40	TGCGTAACCT GTTACACCTA ATGCCACTGC TAATGATGAT GCCATAATTG TCTTTTTTCAT	2340
	AGTAAAAAAT CCTCCAGTAA TAATTGTnAG TTTATGTTTT TAGTAATTAT AtTTTGaATT	2400
45	TGAATGTCGT AGTgCAAGTT TAAATTGTCT TTTATTTCTT TCaACGGTAC TCACTATATC	2460
	ACaaAAAACC AGCCAGTAAA TTACACTTTC TTTACAAAAC ATTACAATAT CAAGTGTTAT	2520
	TTGtAATGTT GAAATATGGC TGTTTTATAC TGTAATGTGA AATATGTGCC CTTTAGAATC	2580
50	CAATCAACCC TTGAAATAGT CTTTAACACA TAAGATTTTT ACTATATTTA GCTCAACTAT	2640
	TACAGCTTTC GTAATATTAC AGATTGTATT TTTGTTACAT AGCTGTAATA TATCTGACAT	2700
55		

	TACACATGTA	TTGATTGCTA	TTATTGTTGT	ATATTCAAAG	TTTTAAAACA	CACATCTTTT	2820
	GTGAATTGTC	TTATCTTTTA	TTAGCGCAAA	TAAACTGCAG	CTCAATTATA	TTGTTCAACT	2880
5	TCATTCTCGC	AATTCACAAT	AACATTAAAT	AATTTTGGT	CTCATATTTT	CAAAAAACAT	2940
	ACTGTTATTA	TCCCATGAAT	TTAAAAATAT	CATTAGTATA	TAAACGAAAC	ACTTTACGAT	3000
	AAATGATATC	TGCAAGCCAA	GCTGTTACAA	ATGGTACAAC	AAAGAACGCT	ACTACAATTA	3060
10	GTAAGACACT	CAACCAAGCA	GAATCAACCT	CCATAAATTT	AAATGCATTA	ATCGGTCCTA	3120
	CCATTCCCTAT	AAAACCAAAT	CCAGCTGACT	CTTTCGTTCC	ATGAATACCT	ACTAATGCTG	3180
15	ATACCAAACC	TGATACAATG	GCTGTCGTTA	ATATTGGTAA	CATAAGAATT	GGATATTTCA	3240
	CCATATTAGG	TATCATCATT	TTAACGCCTC	CAAAGAAGAC	GGATAACGGC	ACCCCTAAAC	3300
	GATTCACTTT	ACTTGTAACA	ATTATCAATA	CTGCTTCAGT	CGCGGAGATA	CCAATTGACG	3360
20	CTGATCCAGC	TGCTAAACCT	GTAATACCTA	TCGCAAAGGC	AATGGCCACA	GTTGATAGTG	3420
	GCGAAATAAT	AATAAGACTA	AATACCATTG	AAATCAAAAT	ACTCATGACA	ATCGGTTGTA	3480
	ATTCTGTAAA	ACCATTAACC	ATATTACCGA	TGGCTGTTGT	AATCATTTTC	GTATACGGCA	3540
25	ATATTAAAAC	ACCAATTGCA	CCTGAAATAC	CGCCAACAAC	TGTTGGGAAT	ACAATCAATG	3600
	CCATACTACC	TACGCGATGT	TGAATAAGTA	AAATGAATAA	CACTGCAATC	GCTGCTGTAA	3660
	TCATTGTATT	AATTAAATCA	CCAATACCCG	TAATCATCCA	AGCACCATT	TTAAACTGCG	3720
30	CTGCACCGCT	TCCTACATAT	GCTGCACTTG	CCACAACAGC	AATTGCTAAT	GGCGATAGGT	3780
	CAAATTTTCAT	GGCAACCAAT	GCACCAATCA	AAGCAGGTAC	TGTAAATTGA	ATTGCAACGA	3840
	CAACGCCTAA	TAACGTTTTA	AAAATCGGAT	GATAATCCAT	AAAGTATTTA	AAAATTTCTC	3900
35	CAAGTATCGC	ATTAGGAACT	AAACCCGCAA	CAATACCTAT	GGCGACACCT	GATAAACTC	3960
	TAAATATAAA	ATCTTTGGGT	GTAATTGTTT	TAATTGATGT	CATAATATCA	TCCTTCCATT	4020
40	TATGTATATA	CATCTGTATG	CAAATAATAA	AGAGCCTTAA	GTTATAAGCT	GCCACTAGCT	4080
	TAAATCTTAA	GATGTGCATG	CCGATGTTGT	TATATTTAGG	CTAGCAGTAT	CATCTATAAC	4140
	TCAAGACTAT	GAAAAATAGT	ATATCACAAA	ATTCTGAATT	TTTAGATAAA	TAAATTGGCA	4200
45	ATTTTTCAAA	CATATTGTTA	CAATACACTT	TTATTTTATC	TTCATTTTTA	AAATCCATTA	4260
	ATACAATAGA	AGAAAGACAT	TCAAATGCTT	ACCAAAAAGG	TACATTATTT	GTTAGGAGCG	4320
	TATCAGCaCT	TACATATCAT	CAACACAATT	GACAAATATA	TAGAAGATAC	TGATAATAAG	4380
50	TGTTAAAACA	ACAGATGTTA	GGTAGTGAAC	AAATGATGGA	AAGTAAATCC	ATAGATCCAA	4440
	GAATCGTTAG	AACCAAACAA	TTGCTTGTCG	ATGCTTTTCT	TAAAATTTCT	AGAGAAAAGA	4500
55							



EP 0 786 519 A2

	TTTACGCTCA	TTTCGCTGAT	AAAGAAGACC	TCCTAGACTA	CACATTATCT	GTAACCATT	4620
	TAAAAGACTT	GAATGATAAT	TTGAGCATT	CTAATGTCAT	TAATGAAAAG	GTTCTGCGTA	4680
5	ATATTTTCAT	TTCAATTGCG	AGTTATATCA	AAGATGCTGC	AAAGTCTTGC	GAATTAAATA	4740
	GTGAAGCATT	TTGCAACAAA	GCACATCAAC	GTATTAATAA	TGAATTAGAA	GATATTTTGT	4800
	CGATTATGTT	AGAAAACAGC	TATCCGGAGC	ATCAACGAGA	TATCATTGTA	AATAGTGCGA	4860
10	GTTTTTTAGC	AGCTGGTATC	TCAGGCTTAG	CATTACATTG	GTTTAACACG	AGTCAAGAGA	4920
	CAGCCGATGT	GTTTATCGAT	CGCAACCTTC	CATTTTAAAT	TCATCATATA	GCACATTTTT	4980
	AATAAACTT	GGTATTTAGT	CATGCATCTT	GAAATCACTA	TGTGACTTAG	GTTCATACTT	5040
15	GTACACACAA	TAAAATTTAA	CGTATTACGA	TTGATTAGCC	GTGTCTAGGA	CATAAATCAA	5100
	CGTCCTATAC	TCTACAATGT	CATATTAGCA	GTCGTAACT	GAATGAAAAT	AAGCTTGTCA	5160
20	TTAAAACATA	TAGATTTTAG	TGACAAGCAT	TTTTGTTTTT	GCGTACTTAA	ACAACACTTC	5220
	AGGCAATATG	TTGTTTAGGC	AACAAATGAT	ATGTGCGTGT	TTATTGGCAA	ACGTACGACA	5280
	TAGTAGTATA	GTATGTCTAA	ACAACATATG	TTGCATAGTT	GATATGCGTT	GTTTAAATAC	5340
25	TAAGATAGGA	GGGATTGACG	TGAGCGAGAC	AGATGAACCT	CAGGGGTTTG	AACGCACGCA	5400
	TAATATATTA	AATATTAATC	AGAGTAGTCT	GGGTGTAGTG	ACATACATTA	CAAATAAATT	5460
	AAAGTCGACG	TTGAAGCAAC	ACATAATAAT	TGCTCGTGGT	AAAAAGCGAA	TCGACTATCG	5520
30	ACTGTCGTAT	AACTTTACAA	TACGTATTAT	GATAATGTAG	AAATCAAGAA	AATCGACTGT	5580
	GAATATACCT	ATGCTATGCC	CATTGCAATT	TTAATAAGAC	ACACGATGTC	ATTCGACAAT	5640
	GCTCATTTCT	TTGCTCAGTT	ACGTCATCCT	GTCTTATAAA	ACAACATTGC	AGACATGTAT	5700
35	ATCAAACGAC	ACTTCAATAA	CATCACTTTG	CCCATCGTAC	TACTAGTAAA	ATCGTGTCTC	5760
	AAATCCCTTA	TTTAAATTCC	AAAAATCTGC	TGGTCAAAAG	ACCGAGAAAC	TAAAAACATT	5820
40	ACTTAATGTG	TTGATAAATT	ACCATATAAA	AATAATCTCA	AAATATATCA	ACACTTGATT	5880
	CTAAGGAGGA	TATGACAATA	TGAAAATTTT	AGATAGAATT	AATGAACTTG	CAAATAAAGA	5940
	AAAAGTACAA	CCACTTACTG	TAGCTGAAAA	ACAAGAACAA	CATGCATTGC	GTCAAGACTA	6000
45	CTTAAGCATG	ATCCGAGGAC	AAGTATTAAC	AACATTTTCC	ACAATAAAAG	TGGTTGATCC	6060
	AATCGGTcAG	GATGTCACAC	CAGATAAAGT	TTATGATCTT	CGCCAACAAT	ACGGTTATAT	6120
	TCaAAATTAA	tATTTGCTCA	CGAGGTATTG	CACTTAAGGT	GCCAAGTAC	CTCATAAACA	6180
50	AAGCCCATAC	TGATTGAAGA	CACTAATGTG	tCsaCCATGG	TGCACATTAC	GCTTCATCTC	6240
	TGTATGGGCT	TTTTATTTAT	TCTTTTGAGA	ATTTCATTTT	AGCAGACCAA	AAAATTAAAA	6300
55							

TGAACGACTG TGCCACCCGC TTCTTTCAC TTAATTCACCA ACTGGTCAAC TTCTTCATTT 6420  
 GTGTTACACAC CTAGAGAAAT CATCACTTCA TTTGGTTCAG TATTAAGGCT TTGCTGACTT 6480  
 5 ACATTTTGAA AATGCTTGTh TTCTATTAAA ATTACGGkTG tTTGACCTAT tTGAATGCCG 6540  
 ACCATTTTAT CTAACATTTG TGGGTTTCTA TTTATTTTAA ATCCTAACGC TTTATAAAAC 6600  
 TGTGCGCTCT TTTCTAAATC TTGCACATGC AAATTAAACC ACATTGATTG AATCATGATT 6660  
 10 GCACCCCAT TACTACTTAT TATAGTTTTG GACTTTAAGC CAATCACTTA ATGATAATCT 6720  
 TGTGGGATTT ATTTAGCCCA TTAATTCAAA GTCTACTTCA TAACCTTTTT CTTCACCA 6780  
 15 TTGCTTTTCT GCAACACCAC TAACAAATTC TCCTTCTATA ACAGTAGATT TACCTGTCAC 6840  
 TTCACTAAAA ATTGTTGCTG CTTCACTTAA TGTAACCTCA TCGGAACCAA TCTCTATTGA 6900  
 TTGATGCGTA AAGCTTTGTG GATGTGCAAA AATATACGAT GCAATTTTAG CTATATCAAT 6960  
 20 AGAAGAAATC ATTGTGAATT TTATATTCGG ATTAATAAAT TCTGGTAATG TAATACGTTT 7020  
 ATCTTCGACT TIAGCAATGC GTAAAAAATT ATCCATAAAG AATGATGGTT TGATAACTGT 7080  
 TGCATTTATA TIAGATTCCA TTAATCTATT TTCTATTTTT GCTAGTACTT CAAAGTGTGG 7140  
 25 GCCAGTTCGA TTTCGATTAA CCCCTCCCGC AGTACTATAC ACAATATGTT GAATATTTTC 7200  
 TTGCTCAGCT ATTTCAATTA TCTTCATACC TTGTCTTAAT TCTTCGCTAA CATCATCTTT 7260  
 AACGATTGGC TGAATACTGT ATAAGCCATA CTTACCTTTC ATCGCTGATT GCAAACCTAAC 7320  
 30 ATTATCACTC AGATCACCTT CArcGATTGA TAAATGCGGA TGTCTATGT CTGAAAGTTT 7380  
 ACGATTAThC TTATTTCTAG TTAATGCACT TACATACCAT CCATCCTCTA ACAACTGTTT 7440  
 TACAACTGCA TTACCTTGCT TCCCTGTTGC GCCTATTACn AAAATATCTT TCAT 7494  
 35

(2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11802 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

AATTTATTTT GCGTCCAC CCCAACTTGC ATTGTCTGTA GAAATTGGGA ATCCAATTTT 60  
 TCTTTGTTGG GGGCCcGGCC CAACTCGCAT TGCCTGTAGA ATTTCTTTTC GAAATTCTCT 120  
 50 GTGTTGGGGC CCCTGACTAG AATTGAAAAA AGCTTATTAC AAGCGCATTT TCGTTCAGTC 180  
 AATTACTGCC AATATAACTT CGTAGATCAT AGAACATTGA TTTATTTCCC AGCCTATTCT 240

55

	AGCAAAGGTA ATAATGATAT TAATAATGTA CAAAAAATAT AAATCAAATC GACATCCTTA	360
	TAAACATCA GAACCACTAA AAACAAAAAA GCACAAAATA AAATTAAATT TAAATAAAC	420
5	GACCACITTT CAAAAAATC TCtTTTCaTa TTTCCACCCC TAATTTTAAT AAGCATTATT	480
	TTATATTCTC TTTTAAGTTT ATTATTCAAA AGGAAAACAG AAATATCTTT CaATATTATT	540
10	ATAAACATTT CAACTACTTT TAAAAACCA CAAAAAATA CTTATTTTAA GTAGATGAGC	600
	ATAAGTGAAC ATAGTTCTTT AGTTATAATA ATTAATTCAA CAAAAGTCG ATTTGTTTTT	660
	GCAATTGGTT TTCATTTCTT CTAAAGATA TTTTCATTAA ATCTGTCAA TCAATAGACG	720
15	CTATATTTTT CACTTATCT CTATATTTAT TTTTAGTAGC TCTTTCTAAA TTTCCCCATT	780
	CCTCTTCTTC GTGAGTTAAT AAATGAAGCA TTGCTCGTTC TTGTATATTT TCAATCATTT	840
	TTAAATTCGG TTTTAAAAA TGCAAATCAT CAAAACAATC TTTCCAACAA TCAACCATAT	900
20	CTCGTTTTAA TTCAATTTCC ACACGCCATA GAAATGTTGA ATCAATTTCA ACATCTGCAT	960
	TATCTTTACG TTCTTGTTTT TATTATAAAT CCGAATAAAC CTATCACTAT TACGCACACC	1020
25	AAAATATTTT GTTTCTGGTT TTACATTACG TCCATAAAAT ATAGTTTCTT TTACCGACTT	1080
	ATCTGACAAT GCATAATAGT CATTTAAATC AAATTCAAAA TCAAAAGCCA AATCTAATCT	1140
	CGTAAACTA ACATCGTCCA AATAACTGAT GATATTTTGT TTTAACCAAA GCACTTCATC	1200
30	ATGCGAAAGC TTATTAGGAT TAAATTCAAC GCGCATAeAC GTCTATTCCA AAGAGTTGCT	1260
	TTTATTTTGT CATATTCAAT ATAACTTTT TCTTTAAGAG CTTTAGCTTT AAAGTTTGTT	1320
	TGTAAATAT CCCAAAGCCG AATTTCAGGA TTAGTACTCA TAAATGTGA AAGTCTCTCT	1380
35	GCGTTAGACA TGCTAAGATT CCCAACATC GTTATAGCGT CAAAAGACAA TTTTGGAATA	1440
	GCTAGTGACA TCCTATGTCG ATTTAACCGG CTATTACCGG ATATTAGAGT ATCCAGTTTT	1500
40	ACAAATGGAT GAAACGAAAT TCAAAACACT AAAAAATATG TTCCACTAAC AGCAAAAAAA	1560
	TACCATTATG TTCCTACTAA AAAACyAAAA ATACTGGAGA ACAAATGTCA GGATATAACT	1620
	TAGGATACTA TGTAATAAAA ATTTACAATA AAAAAACAGG AAAACAAATT TCAAGTAAAA	1680
45	GmATACCCAT ACAAAGAGGA TAAATAAAA AACCTCGAAC TGaAATGATG ATCTTTTCAG	1740
	CTCGAGGTTT AAATATTGGT GCCTTATTTA TATAGATTCG TTATATTATA TTCTCTATTT	1800
	TCATTAAcmT AATCCTTAAA GAGTTTTTAAA TTAATACCTG CTAGATGATT CAAAAATGTT	1860
50	TCATCAACTT TTAAATAATT CAATAATTTT TGTGGTGTCA GTAAATnTCT ATCAAAATAC	1920
	AACTTTAATA AACTATTCAT TTTGACAGGA CGTGACATTT CAATCACGTC GTCTAAAGAT	1980
55	AATACTTTCT CGCTTTAnAC AAAnACAAAA ACTTACCCGA TTAAATCAA GTAAGTTTTA	2040

	TATTTGATAA AAAATCAATA AGTAATTGTG CGCCTTCAAC TTGAATATCT TTTACAAC TG	2160
	GCGCGTCGAT ATACATATCA TACTGACCAC CGCCTACTGC ACGATAATTA TTTACACAAA	2220
5	TTGTATATGT CTGCTTTAAA TCAACTGCGT GACCTTGAAT CATCATATTG CTCACACGTT	2280
	GTCCCTTTGG TCTTCCAACA TGAATGGTAT AACTTACGCC ACCATATATA TCATAATTAA	2340
10	AGTGTTGTGG TTTGGGTTCA AGGAAGTCTG CGCTCACACT AACTTCATCA TTTTTCACGT	2400
	CAAAATATTC TGCTGATCGT TCAATGGCTT CTTTAAGTTT GGCACCACTT ACAGCTAAAA	2460
	CTTTAAATGT ATTTGGAAAT GGGTAATTGT TAATAACATC TCGCATCGTC ACGACTTGCT	2520
15	TGAAACCACT AGCAGAATCA AACAAAGCTG TACAGGCAAC ATCTGCGTCA CTTTTTTCTA	2580
	ATAAAGCGTA ATTCATAAAA TTTGTAAAAG GATGCGGTGC CACACGTGCC TCAAATGCAT	2640
	GATTAATCGT CATATCATAT GGCAATGTAG TAATTTCTGT ATCTAACCAG TCCTCTAACT	2700
20	GCTTTCGTAA ATGTTGGTCA TCTTCATCAA TAGTAAATGT GGAATCATCT ATAACAGGAA	2760
	GTAATTCACA TGATTCAACG GATAGATTTT CATATTCATC AGTACTCAAG ACTACTCTGC	2820
25	CTACAGTTGT ACCTCTCGTA CCAGGTTGAA TCACAGCCGT TTGCTTAAAC CTTTCAGCAA	2880
	TTTGTGATG TTGGTGACCC GTAATAAAGA TATCTATATC TTTAGAAAAC GCTTCTAACA	2940
	TGGCATATCC TTCATTTTCA CCCGTTAATA CTTCGGTCCG CGTACCACTT TCTAAATCCT	3000
30	TTTCAAATCC ACCATGGTAA CAAACCACAA TGATATCTGC ATGTCGCTTC ATTTCAGGTA	3060
	AGTATTGTTG AAGTATTTCA AAAGCACTAT GAAACGTAT GnCnTGAATA TGCTCTGGTT	3120
	GTTCCCAATG GGAATAAAT TGTGTCGTTA AACCTATCAC ACCAACAGTT TGATCTCCAA	3180
35	CCTGAAAATA CTTACACACG TTATCAGTCA ATGTACTATC ATTTTCATAT ATATTAGCGC	3240
	ACAAAACCTGG ATAATTGAGT CTGCGTAAAG TGTCTTTTAA GTATGGTAAT CCATAATTAA	3300
	ATTTCATGATT ACCAAGCGTA CCAAAGTCGA ATGCCATTCT ATTATAAAAA TCAACTAAAG	3360
40	GCTGGCTACT GCCGCTATGC GCGATTAAGT AATTACAAAA TGGTGACCCT TGCAAAAAAT	3420
	CACCATTATC TATTTTAAAA CTTTGGTCAT ACTGCCTTCT GTsTTGTTCT ATAACATGAT	3480
45	TCGCTAGTAA CAATCCCATTA GGTGATATT GATTCTACT CGTAAAATCT GTTGGGAAAA	3540
	TATAACCATG TACGTCACTC ACGACATAAA ATGCTATGTT TGACATCCTC ACTCACTCCT	3600
	TCAATCACAA ACATCTTTCT TATTTCTATT ATATATTTAT TTGAAGTCTG TTGTAATCAA	3660
50	GGTTTTGTCA CCGAGTTTTA AACGAATCTT TGAACCTTCC ATACTTTCAA GTACTTTAGC	3720
	ATTGACCTTA ATTGTGACAT TTCCGTTTTT ATCTGCTTTA ACTGTTGGCA AAGTACTGTA	3780
55	ACCTGGTGGG TTATAATCGT TATCTTTACT TGAAAATTGT CCGATTTGAC GTCCGCCTTC	3840

	TATTGTCATT TCAAATGGCT CATTACAGA AACATTTTGC GGGATATCAA ATGTTACTTT	3960
	TTCGTTCTGA TTTGGTGGTG TATGATCATC TGGTGTGTTT GGCTGAGGAT CTGCGCCTTT	4020
5	TTCGCTGCCA TAACTACCTG CTTTAAATGT TGTGGATCA TACCATTAT AACCACCTCGG	4080
	CGTTGTGAC CATGGCTCTT TTTCAGGCTC AGTTGAACGC TCTGGTCGTT CAAAATCAAG	4140
10	CAACTTAGTC TTTGTATCTA ATGTTAGGCT ACTCGCCTTA AGTGATTTC CATCATTATC	4200
	TTTAGACATC CAAGCCGTTA TATTATTTAA TAGCTTACCG TTGTCTTGTT CTTTAAAAACC	4260
	ATCATATGTT TTCTTCTTTT CTCCATTATC TTCTCTTACA TATTTGGGCG AACTATCTTC	4320
15	CACAAGTGAT GAATCACCGA TAAATGCTGC TTTACCTTTT CCAACTTTAG AAATTGCTAC	4380
	ATAGGGGGCT TCTGCTTTAC CGCCCCATT ATAAATACCT TGATCTACAG CATGTGACCA	4440
	TTTACTTTTC GCTGGCAATT GTTCTGGTGT ATACACAATA CCTTTTGCTT TCTCTGGATT	4500
20	AGTAATTGCT AATGTCGATC CGGCATGCAT AGAGACAGAT TTCACACCTT CAGTAATACC	4560
	GAACTTTTCT TTTGAAGAAA CAATATTGCT CGTATTTTAA TCACCTAGTG CATTATATCG	4620
25	AAAACGTACG CCAAAGTTTG TAGATAACCA ATCTGAACTT TTCACACCTT GCATTGCAGT	4680
	AGAACTTTTT TCTTCTGCAT TCATACCTTT CGACATATCT TCATATGCTC CACGTCGATA	4740
	ACCATTTCATT GCCTCCGATG AATCAATACG ATTTAAATTT CGGTCAGCAT TGTAATGATC	4800
30	TGAAATAAAG ACAACATTGC CACCTTGTTT CACATATTTA ACAATTGCTG CCTGTTCTGA	4860
	TTCTTTGAAA GGAATGTTAG CCTCAGGAAT TACAAATATT TTGGAACCTT TCAAACCTGC	4920
	TTCTGTTATG TTCGAATGAC CATCAATAGC TTTAACGTCA TAACCTTGTT TTTGTATTGA	4980
35	ATCCGCATAA TCTGAAAATG CACCATCACT AACCCAATCT GCAGCACCAG CTGTTTGACC	5040
	ATGAGAACGA TCGAATAATA CCGTTCGCTG TTGCTTTGTA GGTTGCGATT CATGCGTTAT	5100
	AGCTAAAGAT TGCGGTAAAG CACTTAATGA TACCGTTGCA ACAATTGCAG AGACAGTTAA	5160
40	TGACTTATAT ATTTTTTTCA TTTGTGAGG CTCCTTTTAA AATAAATTG TTCTTGAATT	5220
	ATAGGATAAA AATTCGTTGC ATATGAGCAA TTTAACGAAA AATTTACAAA ATCTTATCAA	5280
45	ACTCTTAAAG AAAGTTATTA AAATTCATTT TTATAAAATA CTTTTTAACA TTTAAATGTG	5340
	GTACGCTATA AGTGTAATTT CATTGCATAC ATATTACAG ATTAAGAATG TGAAGGGGAC	5400
	AGTTATCAAA TGAAAAATTT TAAGTGTTTA TTTGTATTAA TGTTAGCAGT CATTGTTTTT	5460
50	GCAGCAGCAT GTGGAAACTC AAGTTCTTTA GATAATCAAA AGAACGCTAG TAATGATTCTG	5520
	GATTCTAAAT CAGGAGGATA CAAACCTAAA GAATTAACCG TTCAATTTGT ACCTTCGCAA	5580
55	AATGCTGGAA CATTAGAAGC TAAAGCAAAA CCATTAGAAA AATTACTATC TAAAGAATTA	5640

	TCTAAAAAAG TTGATGTTGG TTTCTTACCA CCAACGGCAT ACACATTAGC ACATGATCAA	5760
	AAAGCAGCTG ATTTATTATT ACAAGCACAA CGTTTCGGTG TAAAAGAAGA TGGTTCAGCA	5820
5	AGTAAAGAAC TTGTAGATAG TTATAAATCA GAAATTCCTG TAAAAAAGA CTCAAAAATT	5880
	AAAAGCTTGA AAGATTTAAA AGGTAAGAAA ATTGCCTTAC AAGATGTAAC ATCAACTGCT	5940
10	GGATATACAT TCCCCTTGC GATGTTAAAA AACGAAGCAG GTATTAATGC AACTAAAGAT	6000
	ATGAAAATTG TGAATGTTAA AGGTCATGAC CAAGCAGTTA TCTCATTATT AAATGGAGAT	6060
	GTAGATGCTG CGGCTGTATT TAACGATGCA CGTAATACTG TGAAAAAAGA CCAACCAAAT	6120
15	GTATTTAAAG ACACACGAAT TTTAAATTA ACACAAGCTA TTCCGAATGA CACAATTTCT	6180
	GTAAGACCAG ATATGGATAA AGATTTTCAA GAAAAATTGA AAAAAGCTTT TATAGACATT	6240
	GCTAAATCAA AAGAAGGTCA CAAAATTATT AGCGAAGTTT ATTCACATGA AGGATACACA	6300
20	GAAACGAAAG ATTCAAATTT CGACATTGTA AGAGAGTAGC AAAAATTAGT TAAAGATATG	6360
	AAATAATCAT TATTTAACAA ATGAATCATT AGCGAATTTG GTATTAAAAG CTTTCGTTCA	6420
25	ATAGATATAT TCTAGATTAA TATTGAAAAG CTAGGCGCTA AACTGAAACA GATATAGAAA	6480
	GGTGTCGCTG TACATTTGAA ACCATTTGTA CACAGAAACC CAATGTCTAT GATATTTAG	6540
	TTTACCTTGG CTTTTCTTTA TTAAAGAAAAG GTGTCAAACA TGAGTCAAAT CGAATTTAAA	6600
30	AACGTCAGTA AAGTCTATCC TAACGGTCAT GTAGGCTTGA AAAATATTAA CTAAATATT	6660
	GAAAAAGGTG AATTTGCAGT TATTGTCGGA CTATCTGGTG CTGGGAAATC CACGTTATTA	6720
	AGATCTGTAA ATCGTTTGCA TGATATCACG TCAGGTGAAA TTTTCATCCA AGGTAAATCA	6780
35	ATCACTAAAG CCCATGGTAA AGCATTATTA GAAATGCGCC GAAATATAGG TATGATTTTC	6840
	CAACATTTTA ATTTAGTTAA ACGGTCAAGT GTATTACGAA ATGTACTAAG TGGACGTGTA	6900
	GGTTATCACC CTACTTGAA AATGGTATTA GGTTTATTCC CAAAAGAAGA CAAAATTAAG	6960
40	GCAATGGATG CACTAGAACG CGTCAATATC TTAGATAAAT ATAATCAACG CTCTGATGAA	7020
	TTATCAGGTG GCCAACAAACA ACGTATATCT ATTGCACGTG CGCTATGCCA AGAATCTGAA	7080
45	ATTATTCCTG CAGATGAACC AGTTGCTTCA TTAGACCCAT TAACTACGAA ACAGGTTATG	7140
	GATGATTTAA GAAAAATCAA CCAAGAATTA GGCATCACAA TTTTAATTAA TTTACATTTT	7200
	GTTGACTTGG CAAAAGAATA TGGCACACGC ATCATTGGTT TACGTGATGG TGAAGTTGTC	7260
50	TATGATGGTC CTGCATCTGA AGCAACAGAT GACGTATTTA GTGAAATATA TGGACGTACA	7320
	ATTAAAGAAG ATGAAAAGCT AGGAGTGAAC TAACATGCCT TTAGAAATAC CTACAAAGTA	7380
55	TGACTCCCTT TTAAAGAAAA AGGTTTCTTT AAAACGAGT TTTACCTTCA TGTTAATCAT	7440

	AATACCTCAA ATAGGTGATC TATTCAAACA AATGATTCCA CCTGATTTCG AGTATTTACA	7560
	ACAAATTACA ACGCCAATGT TAGATACCAT TCGAATGGcT ATCGTAAGTA CAGTATTAGG	7620
5	TAGCATCGTT TCAATACCAA TTGCGTTATT ATGTGCTAGC AATATCGTTC ATCAAAAGTG	7680
	GATTTCATA CCCTCGCGCT TTATTTTAAA TATAGTTCGT ACTATTCCAG ATTTGTTATT	7740
10	AGCAGCAATC TTTGTGGCTG TATTTGGAAT CGGTCAAATT CCAGGGATAT TAGCACTGTT	7800
	TATTTTAACT ATCTGTATTA TTGGAAAATT ATTATATGAA TCATTGGAAA CGATAGATCC	7860
	AGGTCCAATG GAAGCAATGA CGGCTGTTGG CGCTAATAAA ATAAAATGGA TTGTTTTCGG	7920
15	TGTTGTACCA CAAGCCATAT CGTCATTTAT GTCATACGTA TTATATGCAT TTGAAGTAAA	7980
	TATACGTGCT TCAGCTGTGC TTGGATTAGT CGGCGCTGGC GGTATTGGAT TGTTTTATGA	8040
	TCAAACACTT GGTTTATTTT AATATCCAAA AACAGCAACG ATTATTTTAT TTACTTTAGT	8100
20	TATCGTCGTC GTCATTGATT ACATCAGTAC GAAAGTGAGG GCACATCTCG CATGACACAG	8160
	GAAATAGCAA AATATAATGT TCACACAAAA GCACACAAAC GAAAATTGAT TAAAAGATGG	8220
	CTTATTGCAA TTGTCGTCTT AGCTATTATC ATCTGGGCAT TTGCAGGTGT ACCAAGTTTA	8280
25	GAACTTAAAA GTAAATCATT AGAAATCTTA AAATCCATAT TCAGCGGATT ATTCCATCCT	8340
	GATATCAGCT ATATCTATAT ACCAGATGGC GAAGACTTAT TACGTGGTTT ACTTGAAACC	8400
30	TTTGCGATAG CCGTTGTAGG TACTTTTCATC GCCGCAATTA TCTGTATTCC ATTAGCATTT	8460
	CTAGGTGCAA ATAATATGGT AAAGCTACGC CCAGTTTCAG GTGTTAGCAA ATTTATTTTA	8520
	AGTGTATATC GTGTCTTCCC AGAAATTGTA ATGGCACTTA TATTTATCAA AGCTGTGGC	8580
35	CCAGGTTTAT TTTAGGTGT ATTAGCTTTA GGTATCCATT CCGTAGtATG CTTGGGAAAC	8640
	TTTTAGCTGA AGATATTGAA GGTCTAGATT TCAGTGCTGT AGAATCATT AAGGCCAGTG	8700
	GTGCGAATAA GATTAAAACA CTCGTATTTG CAGTCATACC ACAAATTATG CCTGCCTTTC	8760
40	TATCACTCAT ACTTTATCGC TTTGAACTAA ACTTACGTTT AGCTTCTATA CTGGGGCTAA	8820
	TTGGGGCTGG TGGTATCGGG ACACCACTCA TATTTGCCAT TCAAACACGT TCTTGGGACC	8880
	GTGTAGGTAT TATATTAATC GGTTTAGTAC TAATGGTCGC AATTGTGGAT TTAATTTCCG	8940
45	GTTCAATCCG AAAACGTATT GTTTAACATT AAATCAGGAT ACTCCTAAAT AAGAAGTCCT	9000
	ACCGTCTTAC GTTTCTCTAT TATAATAAAA ACAGCAGTGA AGAAACTAT TGTATAGTT	9060
50	AACTTCACTG CTGTTTTTAT AATATCTAAA TTTATTCTAT TTCAATTCCT TTAAATAACT	9120
	TTTACCGAAC TCTGGTAATG TTACGTTGAA ATTATCTGCT ATAGTTGCAC CGATAGAACT	9180
	GAATGTAGTA TCACTTTCTA GTGCATGACC ACCTTTAAAT TTCGGACTGT ACATAATTAC	9240
55		

EP 0 786 519 A2

	TGTAATAATT ACTAAATCGT CTTCTTTTAA GTTGCTAAAC AGTTCTGGCA AGCGATCATC	9360
	GAAATCTTTA ATTGCTTGTG CATAACCTGG TTTATCACGA CGATGACCGT ATAATGCATC	9420
5	AAAGTCTACT AAGTTTAAGA AGCTAATACC TGTGaaATCT TTCTTAACAA TTTTCATCAA	9480
	TTGATCCATA CCGTCCATGT TACTCTTCGT ACGAACCGCT TCTGTTACAC CTTCACCATC	9540
	ATAAATGTCA TTAATTTTAC CGATGGCAAT AACATCATAA CCACCGTCTT TCAAATGATC	9600
10	TAAGACAGTT TTACCAAAAG GTTTTAACGC ATAGTCATGT CGATTAGATG TACGTGTAAA	9660
	GTTTCCTGGT TCACCAACAT ATGGACGTGC GATAATACGA CCAATTAAAT ATTTAGGGTC	9720
	TTTTGTCAAC TCACGAACCT TTTCAAAAT ATCATATAAC TCTTCTAATG GGATAATGTC	9780
15	TTCATGTGCA GCAATTTGCA ATACTGGGTC TGCACCTGTA TAAACAATTA AGTCACCAGT	9840
	TTTCATTTGG TGCTCGCCCC ACTCATCGAT AATTGCGTA CCCGATGCCG GTTTGTTAGC	9900
20	AACAACTTTA CGACCTGTCA TTTCTTCAAT TTGTGAATT AACTCTTCAG GGAATCCATT	9960
	AGGGTATACT TTAAAAGGTT GCATAATATT TAATCCCATTA ATTTCCAGT GACCAGTCAT	10020
	TGTATCTTTA CCAACTGAAG CTTCACTCAA TTTAGTATAG TATGCTTCTG GTTGTTC AAC	10080
25	TGCATTTACT ACTGGTAATT TATCGATGTT CCCTAGACCT AACTTTTCAA GGTTTGGTAA	10140
	AGTTTGATCG AAACCTTCTA AGGTATGTCT TAAAGTATGT GAACCTTCAT CTTTAAATC	10200
	AGCTGCGTCT GGCGCTTCAC CAATACCTAC TGAATCCATT ACGATTAAAT GTACACGATT	10260
30	AAATGGTCTT GTCATAGCTA TCACTCCCAA AATTTATATA TATTAGTAAT CTGAATCTGC	10320
	TTCTAAACCT TGCATAATTT GAACACCTGC GCTCGACCA ATACGTGTG CACCTGCTTC	10380
35	AACCATTTTA TTGAAATCTT CTAAATTACG TACGCCACCT GATGCTTTTA CTTCTACATC	10440
	AGCACCTACT GTATCTTTCA TTAATTTAAC GTCTTCTGCA GTCGCACCGC CACCTGCAAA	10500
	ACCTGTTGAA GTTTTAACGA AGTCCGCACC AGCCGCTTTT GTTAATTCAC TCGCTTTTAC	10560
40	AATTTGCTCA TGGTCCAACA ATACCGTCTC AATAATCACT TTTACTGTGT GACCTTTGCG	10620
	AGCTTTAACC ACTGCTTCAA TGTCTTGTG TACATCATCA AAACGTCCAT CTTTTAATGC	10680
	GCCGATGTTG ATGACCATGT CAATTTTCATC TGCACCATT TGAATTGCAT CTTCTGTTTC	10740
45	AAATGCTTTC GTTGCAAGTTG TCGACGCACC TAATGGGAAT CCTATTACCG TACAAACGAG	10800
	CACCTCTGAA TCAGCTAGTC GCTCTGCTGC ATATTTAACA TGTGTTGGAT TCACACATAC	10860
	AGATTTAAAA TTGTATGctT TCGCTTCATC GATGATTGTA TCGATTGCG TACGTGTTGA	10920
50	CTCAGGCTTC AATAAAGTGT GATCTATATA TTTCTCAAAT TTCATACTTA CTACTCCTCG	10980
	TGTTATATAA TCTCTTTATT TAATTTTACT ATAAATACGA ATATATCTCG CGAATTTATA	11040
55		



5 ATACTCATTAA AACCTAAAAT AATTAAAATA ATACCGAAAT GTGAACTTAA TGCATCATTG 11160  
 CCTGGGAAAT TTAATGCTTT AAAATCGATT AGAGCCGCAG CAATCGCAAT ACCTACAGAT 11220  
 10 ACCGCCACAT TAATAATTAA ATTATAAAAA CCAATAGCCA CACCTGTCAT ATTAAGATCT 11280  
 ATTGTTTTAA TGGCTTCGTT AAGTAAAGGT GCATACATTA AAGCAAAGCT ACCTGCAAAG 11340  
 AATATCATAG AAATGACGAA GATTGAAATG TGATTACCTA CTGCAAATGC AGGTAAAATC 11400  
 15 AAGCTCAGTG CTATTAATAAT AATTGCTGTG ATAATCGCTT GTTTTGAATT CAGATATTCG 11460  
 CCGATTTTAC CACTTAGTGC ACCAACAATG ACTGCTACTA TATAACCCGG TACTAATAAC 11520  
 AGTGATGTTG TGTCTAGTTG CAGATGATAA ATTTGCTCCA TTATGAATGG GAACGTAAAA 11580  
 ATATAACCCA ATTGGATAGC ATACATTACA AATACTATAA ATAAAAATGA AGCATAACGT 11640  
 TTATTTTGGG AAAATGATTT ATTTACTAAT GGACGTTGCG CATTTTAAAT ATATAGCGCA 11700  
 20 AAAACGATAA TCGCAATTAA GGCACCAATC ATATATAACC AATTAAAGTT CGTAATAAAC 11760  
 AGCATGACTG TTGTAGCAGG GGATCCTCTA GAGTCGAnCC TG 11802

(2) INFORMATION FOR SEQ ID NO: 71:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1196 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

35 CTAAAGAAGA TCGAAACAA GATGTTGATA AACAAGTTCA AGCTTTAATT GACGAAATCG 60  
 ATCAAAATCC AAATCTAACA GATAAGGAAA AACAAGCACT TAAAGATCGT ATTAATCAAA 120  
 TACTTCAACA AGGTCATAAC GACATTAACA ATGCGATGAC AAAAGAAGCA ATTGAACAAG 180  
 40 CAAAAGAACG TTTAGCGCAA gCATTGCAAG ACATCAAAGA TTTAGTGAAA GCTAAAGAAG 240  
 ATGCGAAAAA TGATATTGAT AAACGTGTAC AAGCTTTAAT TGACGAAATC GATCAAAATC 300  
 CAAATCTAAC AGATAAGGAA AAACAAGCAC TTAAAGATCG AATTAATCAA ATACTTCAAC 360  
 45 AAGGTCATAA CGACATTAAC AATGCGCTGA CTAAAGAAGA AATTGAGCAG GCAAAAGCAC 420  
 AACTTGACAA AGCATTGCAA GACATCAAAG ATTTAGTGAA AGCTAAAGAA GATGCGAAAA 480  
 50 ATGCAATAAA AGCCTTAGCT AATGCGAAgc GTGATCAAAT CAATTCAAAT CCAGATTTAA 540  
 CACCTGAGCA AAAAGCAAAA GCGCTCAAAG AAATTGACGA AGCTGAAAAA CGAGCACTAC 600  
 AAAACGTTGA GAATGCTCAA ACTATAGATC AATTAAATCG AGGATTAAAC TTAGGTTTAG 660

55

TTGAAGCAAC ACCTGAGCAA ATCCTAGTTA ATGGTGAAC TATTGTACAT CGTGATGACA 780  
 TCATTACAGA ACAAGATATT CTTGCACACA TAAACTTAAT TGATCAGCTT TCAGCAGAAG 840  
 5 TCATCGATAC ACCATCAACT GCAACGATTT CTGATAGCTT AACAGCAAAA GTTGAAGTTA 900  
 CATTGCTTGA TGGATCAAAA GTGATTGTTA ATGTTCTGT AAAAGTTGTA GAAAAAGAAT 960  
 TGTCAGTAGT CAAACAACAG GCAATTGAAT CAATCGAAAA TCGGCACAA CAAAAGATTA 1020  
 10 ATGAAATCAA TAATAGTGTG ACATTAACAC TGGAACAAAA AGAAGCTGCA ATTGCGAAG 1080  
 TTAATAAGCT TAAACAACAA GCAATTGGAT CATGTTAAC AATGGCACCT GGATGTTCCA 1140  
 15 TTCAGTTGAA GGAAATTTCA ACAACAAGGA ACAAGCGCCG GATTGGAACA ATTTGA 1196

## (2) INFORMATION FOR SEQ ID NO: 72:

## (i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1519 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

CAATCGTTTC AACGCTATTA TCTTTAGACA ACAATTGTAA GCGTGTATGT GCAGTTTCTA 60  
 AACAGTCTAT AATTCGAGTT CTTAATTCAG CTGGATCATC TTAAAAATA AAATCCATCG 120  
 30 CTGCAACTTT GTAGACAAAT GTTAAATAGG TAAGTTCACT GTGACTCGTA ACGAAAATAA 180  
 TGTTACCAAC TGGGTCATGC TTACGAATTT CACTGCCTAA TTTGATACCA TTAATATCAG 240  
 35 TTGAAAGTTG AATATCTAAA AAGTAACAGC CTATGTCATT CATATTTTTA GCTTGCTCAA 300  
 GCACCTCATA AGGATTATCA GTTGCAGAGG CAATTTCCAT AGGCTTTTCT TCTATCATT 360  
 TATAATTTTT AATAATGGTA ACCATGTTTT CTCTTTGTTT TGGATCGTCT TCGCAAATGA 420  
 40 AAATTTTCAT ACATTCACAT CCTTATGGCT AGTTGTTAAT AATTTCAACT TTTTGAATAA 480  
 AGAAACCATT TTCGATAATT GTATCTAATA AGACATTGTC TGCATTATCA GCAATTTCTT 540  
 TTAAAGTTGA TAGACCTAAA CCACGACCTT CACCTTTAGT AGAAAACTT TCTTGAACA 600  
 45 ATTCATGAAT GCGTGGTATA TCATCAGCGC ATTTATTCAT AACAATAAAC GTTACTGAAT 660  
 TTCACTTTC AATAAATGCA ACGCGAATGA TAGGGTCATC AATTTCACTT GATGCCTCAA 720  
 TTGCATTATC AAGAATAATA CCAATACTGC GACTTAAATC GATCATATTC AAGTTAATGC 780  
 50 TACTTACTTC ATCGGTATT TCGATACTAA TCGGAATATT CATTTCTGT GCACGTAAAA 840  
 TTTTCGCAGT AATTAAGCCT TTAATTTTAC GTACTTTAAG ATTCTCGATA CCATTTAATT 900

55

GTAGGCCAGG CATGTCATCT TCTCGAATGT ATTCTGAAAG TGTCGTTAAG ATATTGACAT 1020  
 AATCATGACG GAACTTGCGC ATTTTCGTTGT TGATAGCTTC AATCTTCAAT GTATATTTCAT 1080  
 5 AATAGGTTTC AATTTCTTCT TGATTACGTT TATATTTTCAT CTCTTTAAGG AGAAATTGAG 1140  
 AAATAACAAA TGTTAATATA CTTAAAAATA TAGTGATACC AATAAAAAATA AAAGAATACT 1200  
 GCCTTATTAC TTTAGCTTCA TCCGAGTTTA TTTGTGAATA AAAGAAAAAT AATGAAAAAG 1260  
 10 TAAGCAGTAA GATAGTCGAA ATAACTATTA AAAATCCTTT GTTTAGTATT AGATATGGTG 1320  
 TGCTAATTTT TTTGAGAACT CTATTTATTA TATATGAGAA TAGTATACTA ATAGTCACAT 1380  
 AAATAACAAA AAAGCTAGGG AATATTACAA ATATACTATC AGAAATTTTG GTGGATATAT 1440  
 15 GCATATATAA CTATATACCT GTAGTTAGCA CnGTnATAGG AATAATChGG CGAGGTCCAT 1500  
 AATCCACCAA AATAGAATA 1519

20 (2) INFORMATION FOR SEQ ID NO: 73:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5445 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

30 GTAGGAATCT CTTTGTCTTT TTGGGAGGAC ATTTAATATG AATGTATATT TAGCAGAATT 60  
 CCTAGGAACT GCAATCTTAA TCCTTTTTTG TGGTGGCGTT TGTGCCAATG TCAATTTAAA 120  
 GAGAAGTGCT GCGAATGGTG CTGATTGGAT TGTCATCACA GCTGGATGGG GATTAGCGGT 180  
 35 TACAATGGGT GTGTTTGCTG TCGGTCAATT CTCAGGTGCA CATTTAAACC CAGCGGTGTC 240  
 TTTAGCTCTT GCATTAGACG GAAGTTTGA TTGGTCATTA GTTCCTGGTT ATATTGTTGC 300  
 TCAATGTTA GGTGCAATTG TCGGAGCAAC AATTGTATGG TTAATGTACT TGCCACATTG 360  
 40 GAAAGCGACA GAAGAAGCTG GCGCGAAATT AGGTGTTTTT TCTACAGCAC CGGCTATTAA 420  
 GAATTACTTT GCCAACTTTT TAAGTGAGAT TATCGGAACA ATGGCATTAA CTTTAGGTAT 480  
 45 TTTATTTATC GGTGTAAACA AAATTGCCGA TGGTTTAAAT CCTTTAATTG TCGGAGCATT 540  
 AATTGTTGCA ATCGGATTAA GTTTAGGCGG TGCTACTGGT TATGCAATCA ACCCAGCACG 600  
 TGATTTAGGT CCGAGAATTG CACATGCGAT TTTACCAATA GCTGGTAAAG GTGGTTCAAA 660  
 50 TTGGTCATAT GCAATCGTTC CTATCTTAGG ACCAATTGCC GGTGGTTTAT TAGGTGCAGT 720  
 GGTATACGCT GTATTTTATA AACATACATT TAATATTGGT TGTGCAATTG CrATTGTTGT 780

55

	CGAATCAATT TACTAAAATA AAAAGAAACG TAAATAGCAT AATTTAACAT GTTTGATTCA	900
	TGGATTATGC TATTTTTTCG CCAAAATTTA ACAGATTTTG TACAATGGGT TAGCGATTAT	960
5	TTTTTAATAA AGGAGATACT ACTAATGGAA AAATATATTT TATCTATAGA CCAAGGAACA	1020
	ACAAGCTCAA GAGCGATTTT ATTCAATCAA AAAGGGGAAA TTGCAGGGGT AGCACAAACGT	1080
	GAGTTTAAGC AATATTTTCC ACAATCAGGT TGGGTGTAAC ATGATGCAAA TGAAATTTGG	1140
10	ACATCTGTGT TAGCTGTAAT GACGGAAGTA ATTAATGAAA ATGATGTTAG AGCTGATCAA	1200
	ATTGCAGGTA TCGGTATTAC AAACCAACGT GAAACAACGG TTGTTTGGGA CAAaCATACT	1260
15	GGCCGCCCAA TTTATCACGC AATTGTTTGG CAATCACGTC AAACACAATC AATTTGTTCA	1320
	GAATTAAAAC AACAAAGATA TGAACAAACA TTTAGAGATA AGACAGGATT ACTTTTAGAT	1380
	CCGTATTTTG CAGGTACAAA AGTTAAATGG ATTCTAGACA ATGTTGAAGG TGCACGAGAA	1440
20	AAAGCAGAAA ATGGCGATCT ATTATTTGGA ACGATTGATA CTTGGTTAGT ATGGAAATTA	1500
	TCaGGaAAAg CtGCGCATAT TACTGATTAT TCaAATGCGA GTCGTACATT AATGTTTAAT	1560
	ATCCATGATT TAGAATGGGA CGATGAGTTA TTAGAACTat TACAGTACCT AAAAATATGT	1620
25	TGCCAGAAGT TAAAGCTTCG AGTGAAGTAT ATGGTAAGAC AATTGATTAC CACTTCTATG	1680
	GTCAAGAAGT ACCAATCGCT GGAGTAGCTG GTGATCAACA AGCAGCATTa TTTGGACAAG	1740
	CTTGCTTCGA ACGTGGTGAC GTGAAAAACA CATATGGAAC TGGTGGCTTC ATGTTAATGA	1800
30	ATACAGGTGA CAAAGCGGTT AAATCTGAAA GTGGTTTATT AACACAATT GCTTATGGTA	1860
	TTGATGGAAG AGTAAATTAT GCGCTTGAAG GTTCCATCTT TGTTTCGGGT TCAGCAATCC	1920
35	AATGGTTACG TGATGGATTA AGAATGATTA ATTCAGCACC ACAATCAGAA AGTTATGCCA	1980
	CACGAGTTGA CTCTACTGAG GGTGTTTATG TTGTTCCAGC TTTTGTAGGT TTAGGAACAC	2040
	CATATTTGGA TTCTGAAGCA CGTGGTGCGA TTTTCGGTTT AACACGTGGA ACTGAAAAAG	2100
40	AGCACTTTAT CCGTGCAACT TTAGAATCAC TATGTTACCA AACTCGTGAC GTTATGGAAG	2160
	CAATGTCAAA AGACTCTGGT ATTGATGTCC AAAGTTTACG TGTCGATGGT GGTGCAGTTA	2220
	AAAATAACTT TATTATGCAG TTCCAAGCAG ACATTGTTAA TACTTCTGTT GAAAGACCTG	2280
45	AAATTCAAGA AACTACAGCT TTAGGTGCTG CATTTTTGGC AGGTTTAGCA GTTGGATTCT	2340
	GGGAGAGTAA AGATGATATC GCTAAAAACT GGAAATTAGA AGAAAAATTC GATCCGAAAA	2400
	TGGATGAAGG CGAAAGAGAA AAATTATATA GAGGTTGGAA AAAAGCTGTT GAAGCAACAC	2460
50	AAGTTTTTAA AACAGAATAA ACTTGTAGAT TAGACTTTTG TATAAACATT GTGATACAAT	2520
	CAATTTAAGT TAATATTTGA ATCGAGAAGC GAGAGATTG TTCGAACATG TACAATTGAA	2580

55

	GCATTGTCTA CTTTTAAGAG AGAACATATT AAAAAGAATT TAAGAAATGA TGAATATGAT	2700
	TTAGTAATTAA TTGGTGGCGG TATTACAGGT GCAGGTATTG CACTAGACGC GAGTGAAAGA	2760
5	GGAATGAAAG TTGCATTAGT TGAAATGCAA GACTTTGCAC AAGGAACAAG CTCAAGATCT	2820
	ACAAAATTAG TCCATGGTGG TTTGCGTTAC TTAAAACAAT TCCAAATTGG AGTAGTTGCC	2880
	GAAACTGGTA AAGAACGTGC GATTGTTTAT GAAAATGGGC CTCATGTTAC GACTCCAGAG	2940
10	TGGATGCTTT TACCAATGCA TAAAGGTGGA ACATTGTTGTA AATTCTCAAC ATCAATTGGT	3000
	TTAGGAATGT ATGATCGTTT AGCAGGTGTT AAGAAGTCTG AACGTAAAAA AATGTTATCT	3060
15	AAAAAAGAAA CTTTAGCTAA AGAACCATTA GTTAAAAAAG AAGGTCTAAA AGGCGGCGGT	3120
	TACTATGTTG AATATCGTAC TGACGATGCG CGTTTAACTA TTGAAGTTAT GAAGCGTGCT	3180
	GCTGAAAAAG GCGCAGAAAT TATCAACTAT ACTAAATCTG AACACTTCAC TTATGATAAA	3240
20	AATCAACAAG TAAATGGTGT TAAAGTTATA GATAAATTAA CTAATGAAAA TTATACAATT	3300
	AAGGCTAAAA AAGTGGTTAA TGCAGCAGGT CCATGGGTTG ATGATGTTAG AAGTGGTGAT	3360
	TATGCACGCA ATAATAAAAA ATTACGTTTA ACTAAAGGTG TACATGTTGT TATTGATCAA	3420
25	TCAAAATTCC CATTAGGTCA AGCAGTATAC TTTGATACTG AAAAAGATGG AAGAATGATT	3480
	TTTGCAATTC CACGTGAAGG AAAAGCGTAT GTAGGTACTA CAGATACATT CTATGACAAT	3540
	ATCAAACTTT CACCATTAAC TACACAAGAA GACAGAGACT ATTTAATCGA TCGGATTAAT	3600
30	TACATGTTCC CTAGTGTTAA TGTTACAGAT GAAGATATTG AATCAACATG GGCAGGAATT	3660
	AGACCATTAA TTTACGAAGA AGGCAAAGAC CCTTCTGAAA TCTCTCGTAA GGATGAAATT	3720
35	TGGGAAGGTA AATCAGGTTT ATTAATATT GCAGGTGGTA AATTACAGG CTATCGTCAC	3780
	ATGGCTCAAG ACATTGTTGA TTTAGTATCT AAACGCTTGA AAAAAGACTA CGGTTTAAACA	3840
	TTTAGTCCAT GTAATACAAA AGGTCTGGCA ATTTTCAGGTG GCGATGTAGG TGGTAGCAAG	3900
40	AACTTTGATG CGTTTGTA GAACAAAAGTA GATGTAGCTA AAGGATTCGG CATTGATGAA	3960
	GATGTTGCAA GACGTTTAGC ATCTAAATAT GGTTCAAATG TTGATGAATT GTTCAACATT	4020
	GCGCAAACAT CTCAATACCA TGATAGCAAG TTACCATTAG AAATTTATGT AGAACTTGTT	4080
45	TATAGTATTC AACAAGAAAT GGTATACAAA CCTAACGATT TCTTAGTTCTG TCGTTCTGGT	4140
	AAAATGTATT TCAATATTAA AGATGTATTA GATTATAAAG ATGCTGTCAT CGATATTATG	4200
	GCAGATATGC TTGATTACTC TCCAGCTCAA ATTGAAGCAT AACTGAAGA AGTTGAGCAA	4260
50	GCAATTAAAG AAGCGCAACA TGGaATAAAT CAACCAGCAG TTAAAGAATA AtTAATTGT	4320
	ACAATCATAA ACTGGTGTCC TGTTTTAAGG GCATCAGTTT TTTTATACGA GATACATTAG	4380
55		

GTTATTAAAG GTGTGAGATG ATGACTGAAA AACAAITTTAA ATTAAGTGTG CAAGATAATA 4500  
 CGAATATTGA AGTTAAAGTG AATTTTACAG ATGTAGATTC AAAAGGAATT ATTCATATAT 4560  
 5 TTCATGGTAT GGCTGAACAT ATGGAACGTT ACGATAAATT AGCACATGCA CTTTCAAAGC 4620  
 ATGGCTTCGA TGTGATACGT CATAATCATC GAGGACATGG TATTAATATT GATGAATCAA 4680  
 CAAGAGGGCA TTACGATGAT ATGAAACGAG TTATCGGTGA TGCCTTTGAA GTAGCGCAAA 4740  
 10 CAGTGAGAGG CAATGTTGAT AAACCATACA TTATAATCGG ACATTCAATG GGATCCGTTA 4800  
 TAGCTAGATT GTTTGTAGAA ACATATCCGC AATATGTTGA TGGTCTAATT TTAAGTGGTA 4860  
 CTGGTATGTA TTCATTATGG AAAGGTTTAC CAACCGTTAA AGTGTTACAA CTGATTACAA 4920  
 15 AAATTTATGG TGCTGAGAAA CGAGTTGAAT GGGTTAACCA GTTAGTATCA AATAGTTTTA 4980  
 ATAAAAAnNAT ACGTCCATTA CGTACACAAA GTGATTGGAT TTCTAGTAAT CCAATTGAAG 5040  
 20 TAGATAaCTT TATTAAAGAT CCATATAGTG GaTTTAAATGT GTCAAATCAA TTATTATATC 5100  
 AAACAGCCTA TTATATGCTA CATACATCAC AATTAAAAAA TATGAAAATG TTAAaTCATG 5160  
 CCATGCCTAT ATTATTAGTT TCAGGATATG ACGATCCTTT AGGTGATTAT GGTAAAGGGA 5220  
 25 TTTTAAAAATT GGCGAATATA TATAGAAACG CTGGCATnAA AAATGTTAAA GTGAATCTTT 5280  
 ATCATCATAA ACGTCATGAA GTGTTATTTG AAAAnGATCA TGACnAAATT TGGGAAGACT 5340  
 TGTTTAAATG GTTGAATCAA TTTTATAAAA AATAAAGAAA GTGGAATTAA ATATGAATAA 5400  
 30 AAATAAGCCT TTTATTGTAG TAATTGTGGG GCCAACTGCT TGCAG 5445

## (2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 2569 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TGGCTTGAAC TACGCCAATA AGTCCCCCTA GTACAAGAAT GAATACCATG ATATCGACCG 60  
 45 CTTCTATCGT ACCTTCAACC ATGCTACTTG TTATTTGTTT TGGTCCAGCT GGATGTTGCT 120  
 TTAATCTTTC ATAAGTATTC GGAATTGATA CCGGCTTATT AATTGCACCT GATTTAAATT 180  
 GTTCAATCTT AATTTTAAAC CCCATTTTGT CTAGTTCCTG TTGCGTACCC GGAACCTTTT 240  
 50 TCACTTGGTT ATGAGGGTTA ACTATCTTTA GTTCTTGGGA TGAAGGTTCT TAAGAAAGTT 300  
 TAGAATATGC ACCAGCAGGA ATAACCCATG TTGCTATAAC TGCAACAACC GTTAAATGA 360

55

	TAATTGTATT TTCCACGGTT TCATCTCCTT CGACATTTAA CCTAGCATTT CTACCTTAAA	480
	GATTTTATAA ATATAAATTA AGAAAGTGCA CCCC GCATCA AAATAGAGGC ATTATTTTCA	540
5	GGGGGTGCAC ATAAATAATA AAAATCATGC ATTTGACATA TAGTAATTGA AAAGCGTTTC	600
	AATTCAATTA CTTTTTAATC ACAGTACCTA CTTTACCCTC TAAGGCAGCA TCTAATTCAT	660
	TTAATGATGT TATAAGCACA CTTCTTTTGG GATTGTTTTT AATAAATGAT ATGGCTGCTT	720
10	CAATTTTGG TAACATACTT CTTTTGCAA ATTGATTTTC GTCTATATAT CGTTTTAATT	780
	CATCAACATT TGTGTTTTT AAAGGCTGTT GGTTTTCACT GTTAAAATTA ATATATACAT	840
15	AATCAATTGC TGTTAAAATA ATCAATTGAT CGCATTGAAT ATTAGCACCC AACACGCAC	900
	TTGTTTTATC TTTGTCTATA ACTGCATCAA TACCTTTAAA ACCATCATGT TGCTCTCTAA	960
	TTACTGGTAT ACCTCCACCA CCAGCAGCAA TAACGAGTGT ATCATTTTTTA ATAAGTGTTT	1020
20	TAATACTCTC TAATCAATA ATAGAGATGG GTTGTGGTGA AGGAACAACG CGTCTATATC	1080
	CTCTTCCAGC ATCTTCAACA AATATAAATC CTTTTCTTT TTGAATTTGT TCAGCTTCTT	1140
	CTTTGTTGTA AAATAACCCA ATTGGTTTTG AAGGATTGTT AAATGCCGGA TCATTTTCAT	1200
25	CAACTTCAAC TTGTGCTACT AGTGTTACCA CTGTTTATC CATTCCAATA GAATGCAATT	1260
	CATTTTGTA GCTTCTTGT AATTGATAGC CGATGTAAGC TTGACTCATT GCGCCACATT	1320
	CAGCAAATGG AAATGCCGGA CCTTGGTTAT GTTCTGCAGC ATAGTTAAGT CCCAAATTAA	1380
30	TGCTTCCAAC CTGTGGTCCA TTACCATGAC TAATAACAAT CTCATGTCCT TTTGTnATTA	1440
	AyCCTACTAA TGATTtCGCA GTATTTTAA CAAGCTCGAG TtGgTyCTTG aGGTGATTTn	1500
35	CCTAAAGCAT TACCACCTAA TGCTACTACT ATTTTCGCCA TCATATTCAC TTCCTTATAT	1560
	CATTTAAAT TCACCCAATG TAGCAACCAT GaCTGCTTTG ATTGTATGCA TTCTGTTCTC	1620
	AGCTTCTTGG AATACAACCTG AAGCTTTACT TTCGAATACT TCATCTGTAA CTTCCATTTT	1680
40	TCGAATACCA TATTTTCAA AAATTTGTTG ACCTATTTTC GTATCAGCAT TATGGAAAGA	1740
	TGGTAAGCAA TGCTCAAAAA TAACATTTGG ATTACCAGTT TTATCCATTA TTTCTTTATT	1800
	TACTTGATAT GGTTCaATA ATTCAAGTCG TTCTTTCCAT ACTTCATCAG GTTCACCCAT	1860
45	TGATACCCAA ACATCAGTGT AAATTACATC CGAACCTTTT ACaCCTTGGT CaATATCATC	1920
	TGTGATTAAT ATGTTGCCaC CATTTCaGC GGCAATATTT TTACAGCGAT TTAATAATTC	1980
	ATCTGTTGGA TTTAATTCTT TTGGACAAAC TAAATGGAAG TTCATACCCA TAATGGCAGC	2040
50	ACCTTGCAAT AATGCATTG CAACGTTATT ACGACCATCT CCAACATATG TAAAGTTAAT	2100
	ATCTGCATAA TCTTTTTTTA AGACTTCTTT TGCTGTTAAG AAATCAGCAA GAACTTGAGT	2160
55		

TTCTACTGTT CTTTGTGAAA AACCACGGTA TTCAATGCCA TCATACATTC CACCAAGCAC 2280  
 ACGTGCAGTA TCTTTAGTTG TTTCTTTTTT ACCCATTTGT GATCCAGTTG GGCCTAAATA 2340  
 5 AGTTACATTT GCACCTTGAT CATGCGCTGC AACTTCAAAT GCACATCGCG TTCTTGTA GA 2400  
 ATCTTTTTCA AATAACAGTG CAATATTTT ATTTTAAAC ATAGGCTTTT CAGTGCCAA 2460  
 ATATTTAGCA CGTTTTAAAT CCTCGGAGAG TGTTAATAAG GTTCTACCTC TTGTCGTGAA 2520  
 10 AAGTCTAATA AAGTTAAAAA ACTTCTGTTT CGTAAATTTT TCATTAAaA 2569

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 1273 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CCTGGAACCA TCCaATCGtG CaaATCtTGa AAGaGAATAC GCAACAACAA TTAAATGTAT 60  
 25 TGGAACACTA TATTCCAAAT GACCATCCAG CACTCGTTGA ATTAAAAATA TGGGAACGTT 120  
 GGTTACATAA ACAAGGTTAC AAAGACATCC ATTTAGATAT TACTGCGCAC CACCTAGATC 180  
 CTATTACACA GGTTTATTTA TTCAATGTCA TTTTGCTGAA AATGAATCTC GAGTTTAAAC 240  
 30 AGGTGGTTAT TACAAAGGAA GCATCGAAGG GTTTGGATTA GGATTAAACAC TTTAAGTAAG 300  
 GGAGTATGCA CAATGTTAAG AATCGCCATA GCCAAAGGAC GTCTAATGGA TAGTTTAATT 360  
 AACTATTTAG ATGTAATTGA ATATACGACA TTATCAGAAA CATTAAAAAA TAGAGAACGC 420  
 35 CAATTATTAT TAAGTGTAGA TAATATTGAA TGCATTTTAG TAAAAGGAAG TGACGTGCCA 480  
 ATCTATGTGG AACAAGGAAT GGCAGACATA GGCATTGTTG GTAGCGACAT ATTAGATGAG 540  
 40 CGCCAATATA ATGTTAATAA TTTGTTGAAT ATGCCTTTTG GAGCATGTCA TTTTGCGGTT 600  
 GCAGCGAAAC CTGAAACGAC CAATTATCGT AAAATCGCAA CGAGTTATGT TCATACTGCT 660  
 GAAACATATT TTAAATCAAA AGGTATTGAT GTCGAATTGA TTAAATTGAA TGGCTCTGTT 720  
 45 GAATTGGCCT GTGTTGTAGA TATGGTAGAC GGAATTGTCG ACATCGTTCA AACAGGTACT 780  
 ACGCTAAAAG CGAACGGACT GGTGAAAAG CAACATATTA GTGATATCAA TGCAAGATTA 840  
 ATAACATAA AAGCAGCTTA TTTTAAAAAA TCACAATTAA TAGAGCAATT TATTCGCTCT 900  
 50 TTGGAGGTGT CTATTGCCAA TGCTTAATGC ACAACAATT TTAAATCAAT TTTCATTAGA 960  
 AGCACCATTA GATGAGTCAT TGTATCCaAT TATTCGCGAT ATTTGTCAGG AAGTTAAAGT 1020

55



TTTAGaAATT AGTCATGAmC AAATTAAAGC AGCATTTGAC ACATTAGATG AAAAAACAAA 1140  
 ACAAGCATTa CAACAAAGTT ATGAAAGAAT TAnAGCATAT CAaGAAaGTA TtaAACAGaC 1200  
 5 GaATCAACAG TTAGAAGaAT CAGTGGaGTG tTrTGaAATA TACCATCCmC taGaAAGTGT 1260  
 CCGTATTTAT GTG 1273

## (2) INFORMATION FOR SEQ ID NO: 76:

10

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1308 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

20 GTTGATAAAT TAAAAATGTT TTTATCAGAT ATTCAAAGTT ACCAACAATA TAGTAAAGAT 60  
 CATCCGGTGT ATCAGTTAAT TGATAAATTT TATAATGATC ATTATGTTAT TCAATACTTT 120  
 AGTGGACTTA TTGGTGGACG TGGACGACGT GCAAATCTTT ATGGTTTATT TAATAAAGCT 180  
 25 ATCGAGTTTG AGAATTCAAG TTTTAGAGGT TTATATCAAT TTATTCGTTT TATCGATGAA 240  
 TTGATTGAAA GAGGCAAAGA TTTTGGTGAG GAAAATGTAG TTGGTCCAAA CGATAATGTC 300  
 GTTAGAATGA TGACAATTCA TAGTAGTAAA GGTCTAGAGT TTCCATTGT CATTTATTCT 360  
 30 GGATTGTCAA AAGATTTTAA TAAACGTGAT TTGAAACAAC CAGTTATTTT AAATCAGCAA 420  
 TTTGGTCTCG GAATGGATTA TTTTGATGTG GATAAAGAAA TGGCATTTC ATCTTTAGCT 480  
 TCGGTTGCAT ATAGAGCTGT TGCCGAaAAA GAACTTGTGT CAGAAGAAAT GCGATTAGTC 540  
 35 TATGTAGCAT TAACAAGAGC GAAAGAACAA CTTTATTTAA TTGGTAGAGT GAAAAATGAT 600  
 AAATCATTAC TAGAACTAGA GCAATTGTCT ATTTCTGGTG AGCACATTGC TGTCAATGAA 660  
 CGATTAACTT CACCAAATCC GTTCCATCTT ATTTATAGTA TTTTATCTAA ACATCAATCT 720  
 40 GCGTCAATTC CAGATGATTT AAAATTTGAA AAAGATATAG CACAAATTGA AGATAGTAGT 780  
 CGTCCGAATG TAAATATTTT AATTGTGTAC TTTGAAGATG TGTCTACAGA AACCATTTTA 840  
 45 GATAATGATG AATATCGTTC GGTTAATCAA TTAGAACTA TGCAAAATGG TAATGAAGAT 900  
 GTTAAAGCAC AAATTAAACA CCAACTTGAT TATCGATATC CATATGTAAA TGATACTAAA 960  
 AAGCCCTCAA AACAATCTGT TTCTGAATTG AAAAGACAAT ATGAAACAGA AGAAAGTGGC 1020  
 50 ACAAGTTACG AACGAGTAAG GCAATATCGT ATCGGTTTTT CAACGTATGA ACGACCTAAA 1080  
 TTTCTAAGTG AACAAGGTAA ACGAAAAGCG AATGAAATTG GTACGTTAAT GCATACAGTG 1140

55

GATGGATTAA TCGATAAACA TATTATCGAA GCAGATGCGA AAAAGATAT CCGTATGGAT 1260  
 GAAATAATGA CATTATCAA TAGTGATTAT ATTCGATATT GCTGAAGC 1308

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1431 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

GATGCCATTn ATnnGTATGC AAGAAGTTGT TCCGGGTTCA GGTGGATTaC CAGTTGGTAC 60  
 TGGTGGTAAG ACGTTACTAA TGCTTTCAGG CCGTATAGAC TCACCAGTTG CTGGGATGGA 120  
 AGTGATGAGA CGTGGCGTAA CAATTGAAGC GATTCATTTC CATAGTCCAC CATTTACAAG 180  
 TGATCAAGCA AAAGAAAAAG TTATTGAATT GACACGTATT TTAGCTGAAC GTGTTGGACC 240  
 AATTAAATTG CATATTGTAC CATTTACAGA ATTGCAAAA CAGGTAAATA AAGTTGTACA 300  
 TCCAAGATAT ACAATGACTT CAACGAGACG TATGATGATG CGTGTGCTG ATAAATTAGT 360  
 ACATCAAATA GGGGCTTTAG CTATTGTAAA TGGTGAAAAC CTAGGGCAGG TAGCCAGTCA 420  
 AACACTTCAT AGCATGTATG CAATTAATAA TGTAACCTCT ACTCCTGTAT TACGTCCTTT 480  
 ATTAACCTTAC GATAAAGAAG AAATTATTAT TAAATCGAAA GAAATTGGTA CATTTGAAAC 540  
 ATCTATTCAA CCATTTGAAG ATTGTTGTAC AATTTTCACC CCTAAAATC CAGTAACCGA 600  
 ACCAACTTT GATAAGGTAG TCCAATATGA AAGTGTCTTT GATTTTGAAG AGATGATTAA 660  
 TCGTGCTGTT GAAAATATTG AAACACTTGA AATAACTAGT GATTATAAAA CTATTAAAGA 720  
 ACAGCAAACA AACCAATTAA TAAACGACTT TTTATAAATA AAATCCTAGA GTAAATTTAA 780  
 ACATAAGGGG ATGTTAAACT ATGGATTGA ACTTAACGAT GATTATAATC ATAATTTTAT 840  
 TTGTTTTTAT CGCGGCGTTT ATAGATTCCG TTGTAGGGG TGGCGGTTTA ATTTCTACGC 900  
 CAGCATTATT AGCAATCGGT CTACCACCAT CTGTGGCTTT AGGTACAAAT AAATTGGCAA 960  
 GTTCGTTTGG TTCTTTAACT AGTACGATAA AGTTTATAAG GTCCGGTAAA GTGGACTTAT 1020  
 ATGTTGTTGC CAAATTATTT GGTTTTGTAT TTTTGGCATC TGCATGTGGC GCATATATTG 1080  
 CAACGATGGT TCCGTCACAA ATATTGAAAC CTTTAATCAT CATTGCACTT TCGTCGGTGT 1140  
 TTATATTAC ATTACTTAAA AAAGATTGGG GCAATACACG CACGTTTACT CAATTACAT 1200  
 TTAAGAAAGC CATAATATTT GCAGCACTTT TTATATTAAT CGGCTTTTAT GATGGATTTG 1260

TAAGTGCAGC AGGAAATGCT AAAGTTTTGA ACTTTGCTTC TAATATAGGT GCGCTTGTAT 1380

TATTTATGGT ATTAGGACAA GTAGATTATG TAATAGGTTT AATTATGGCT A 1431

5

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4403 base pairs

(B) TYPE: nucleic acid

10

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

AATATTATTT TAAATTCAAT ATTTATTGGT GCATTTATTT TAAACTTATT ATTGCGCTTT 60

ACCATTATTT TCATGGAAAG ACGTTCTGCC AATTCTATCT GGGCTTGGTT ACTAGTCTTA 120

20

GTTTTCTTGC CTTTATTCGG CTCATTTTA TACTTACTAT TAGGACGACA AATTCAACGT 180

GACCAAATTT TCAAAATTGA TAAGGAAGAT AAAAAAGGAT TAGAGTTAAT CGTTGATGAG 240

CAATTAGCTG CTTTAAAAAA TGAAACTTTT TCAAATTCCA ATTATCAAAT TGTAAATTT 300

25

AAAGAAATGA TTCAAATGTT GTTATATAAT AACGCAGCAT TTTTAACAAC AGACAACGAT 360

TTAarrrrtAT ACACAGACGG CCAAGAAAAA TTGATGACC TAATACAAGA CATCCGTAAT 420

30

GCTACTGATT ATATTCATTT TCAGTACTAT ATTATTCAAA ATGATGAATT AGGTCGTACC 480

ATTTTAAATG AACTTGGTAA AAAAGCGGAA CAAGGTGTAG AAGTTAAAAT TCTTTATGAT 540

GACATGGGTT CTCGTGGACT GCGTAAAAAA GGCTTACGCC CGTTTCGCAA TAAAGGTGGA 600

35

CATGCTGAAG CATTTTTCCC ATCAAAATTA CCTTTAATTA ACTTGCGTAT GAACAATCGA 660

AACCATCGAA AAATTGTTGT AATAGATGGG CAAATTGGAT ATGTTGGTGG TTTTAATGTT 720

GGTGATGAGT ACTTAGGTAA ATCAAAAAAA TTCGGCTATT GCGGAGATAC GCATTTACGA 780

40

ATTGTCGGGG ATGCAGTGAA TGCATTGCAA TTACGATTTA TTCTAGATTG GAATTCACAA 840

GCCACACGTG ACCACATCTC CTATGATGAT CGTTATTTCC CAGATGTAAA TTCTGGTGGA 900

ACAATTGGCG TTCAAATAGC TTCTAGTGGT CTGACGAAG AATGGGAACA GATTAAATAC 960

45

GGCTATTTGA AAATGATTTC ATCTGCTAAA AAATCGATTT ATATTCAATC TCCCTATTTT 1020

ATACCTGATC AAGCCTTTTT AGATTCTATT AAAATTGCGG CATTAGGTGG TGTGATGTC 1080

AATATCATGA TTCCTAATAA ACCTGACCAT CCGTTTGTTC TTTGGGCTAC TTAAAAAAT 1140

50

GCAGCATCCT TATTAGATGC CCGTGTTAAA GTATTTCACT ACGACAATGG CTTTTTACAC 1200

TCAAAAACAC TTGTTATAGA TGATGAAATT GCAAGTGTGG GAACAGCTAA TATGGACCAT 1260

55

	AAATTAAAC AAGCTTTTAT AGATGATTTA GCAGTATCTT CTGAATTAAC AAAAGCACGT	1380
	TATGCTAAGC GAAGTCTTTG GATTAAATTT AAAGAAGGTA TTTCACAATT ATTGTCACCT	1440
5	ATCTTATAAA ATAGAAATAT GAGGAGTGTA aCTTTAATGC AACAATCAGA CGTCATTAGT	1500
	GCTGCCAAAA AATATATGGA ATCTATTCAT CAAAATGATT ATACAGGCCA TGATATTGCG	1560
10	CATGTATATC GTGTCACTGC TTTAGCTAAA TCAATCGCTG AAAATGAAGG TGTTAATGAT	1620
	ACTTTAGTCA TTGAACTCGC ATGTTTGCTT CATGATACCG TTGACGAAAA AGTTGTAGAT	1680
	GCTAACAAAC AATATGTTGA ATTGAAGTCA TTTTATCTT CTTTATCACT ATCAACCGAA	1740
15	GATCAAGAGC ACATTTTATT TATTATTAAT AATATGAGCT ATCGCAATGG CAAAAATGAT	1800
	CATGTCACTT TATCTTTAGA AGGTCAAATT GTCAGGGATG CAGATCGTCT TGATGCTATA	1860
	GGCGCTATAG GTGTTGCACG AACATTTCAA TTTGCAGGAC ACTTTGGTGA ACCTATGTGG	1920
20	ACAGAACATA TGTCAC TAGA TAAGATTAAT GATGATTTAG TTGAACAGTT GCCACCATCT	1980
	GCAATTAAAC ATTTCTTTGA AAAATTACTT AAGTTAGAAT CTTTAATGCA TACAGATACG	2040
	GCGAAGATGA TTGCTAAAGA ACGTCACGAC TTTATGATGA TGTACTTGAA ACAGTTTTTT	2100
25	ACGGAATGGA ATTGTCACGA CTAGACATTG AAGTTGTAGT ATGATGATGC GATGTAATGG	2160
	CGTGTGTGTTG TGGAAGCTTG GTGTCATGCC ATGTTACTTT GATGTGTTGT TGTGGGAGCT	2220
	TGGTGACATG TCATGCTACT TTGATGTGCT GGTACCACGA TCGCTCTGA TGTAGTGCTA	2280
30	TGATGTGGCA TTGCGGTGTT ATGGTGTTAT AGACAGGTTT GCGGTGATG CCATGTTACT	2340
	TTGATGTGCT GGTACCACGA TGCGACTTGA TGTAGTGCTA TGATGTGGCA TTGCGGTGTT	2400
35	ATGGTGTTAT AGACCGGTTT GATGTTGATG CCATGTTACT TTGATGTGCT GGTGCTACGA	2460
	TGCGACTTGA TGTAGTGCTA TGATGTGGCG TTGCGCTGTT ATGGTGTTAT AGCCAGGTTT	2520
	GGTGTTGATG TCATGCCGTT ACGATTCTAT GATATGTTGT TGGGACGTTG CAATGTGTAT	2580
40	TATGCCGTTG TGACGTTATT ATTTACACT GTTACATGTA TAAGTGAATT GCTGTGAAA	2640
	TTTGCGACAT ATACTGCTAC ACTGATGAAT CATTGTGTCA AGATGACATT GCGATGAAGA	2700
	ATGACAACTC TGTTATTAAC CACTTTTTTAC ATACTGAAAA CTCGTTAATA TTATTTCAAA	2760
45	TAAAAACAGC AGTAGGATGA CTTTCACATT TGAAATCATC TTAGTGCTGT TTCTATTAT	2820
	CACATATTGT ATAATGTGAC ACTAAGTTTC GCTATTGAAG CGAAAAATAA TGTGCGCCCT	2880
	ATAAAGTTAA AATTATCTTC AACTTTTAGG GTGCACATTA TTTGGACTTG CTAAGTTAT	2940
50	TTCTTTTCT TTTTAGACAC AACTTGTGTG TTTTTCCTT TTTTATTGct GCCGCCGTTG	3000
	TGCTCTCTTT CATACGCTTC AATGAAAGGT TGTACTTCTT TTTTAGCGAC TTTTTCATAA	3060
55		

CCAAGTGCTG ATGCTGAGCT TAATGAAATC CAGATAATCA TAATTGGTGA AATGACCATC 3180  
 ATCATGTAAC CCATTGACG TTGTTTCGTCT GGCATCGTTT TACTTGATAC ATATGCTTGG 3240  
 5 ATAAAGTATA AAACACCGGC AATAATTGTA ATCCAAATAT CAGGACGTCC TAAATCGAAC 3300  
 CATAAGAAGT GTGGATATTT AAACAAACCA TCTACAAGTT GGTCTTTAAG TACAAAGTAT 3360  
 AATCCCATGA TGATTGGTAA TTGGATTAGC ATTGGTAAAC AACCCAACAT ACTCTTAATC 3420  
 10 GGGTTCATGT CATACTTTTT ATATACTTGC ATTAATTCTT GGTTCGAGC CATTTTTTCT 3480  
 TCTTGTGTAC GCGnCaCGTT cACTTTTTCT TGAATTTTTT CAACTTCTGG CTTTGCAACT 3540  
 TTCATTTTTT GACGCAATCAT ATGACTATTT TTATAGTTTG ACAACATGAA TGGTAATAAA 3600  
 15 ATAATACGAA TTACCAATAC AAGGATAATA ATAGCTAAAC CATAATTGTC GTTTAATAAG 3660  
 TTATTCCCA ACCAATCCAA TACATTTTTT ATTGGATCTA CGAATGTATT GTAGAAAAAy 3720  
 20 cwCtACGTTT TTCAGGTTTA GAATAGTCAC AACCAGCCAA AAAGACCATA ATACCTAAAA 3780  
 ATAATGGTAG TAACGCTTTT TTCTTCATTT TTCCACCTCT ATCATTATAT TCACATAGGA 3840  
 TTTATTCTAT CACATTAATG AGTACGTATG AAACAATAAG TGGAAAAATT TAACTAATTA 3900  
 25 TTAAAAAAT CTTTGAATCG ATTAACAGTC TTTCAATAT TTTCACTTT AGAAATGGCT 3960  
 GAAATGACTG AAATTCCATT GGCACCTGCT TCTACAATCG GCGCCACATT ATTAGTATTG 4020  
 ATACCGCCAA TAGCTACAAT CGGTAGTTGC GGATTCATTT CTTTAAACGT TGCAATCATT 4080  
 30 TCTGGACCTA CTGGTATATG CCGTCATGC TTCGACGGCG TAGGATAGAT TGGTCCAACA 4140  
 CCTATATAAT CmACATGAGT TAAATCAGAT TTTGCATACT CATCTAAATC ACTAATACTA 4200  
 AGTCCAATAA TTTTATCAGT GAAATATTGT GCTATCTCTT TGACTTTCGC ATCATCTTGA 4260  
 35 CCGACATGTA TACCATCCGC GTTAATTTCT TTTGCCAAGG ATACATCATC ATTAACGATA 4320  
 AAAGGCACAT CATATTGATG ACAGAGATGC TGTAATTCTT TAGCTAATAC AAGTTTATCG 4380  
 40 TTTCCTTTTA AAGCTGATTC ACC 4403

## (2) INFORMATION FOR SEQ ID NO: 79:

## (i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1808 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

50 TGGAnCCAAT ATTAGAAATG ATTAAACAT TAACAGGTAT TAATAGTCCT TCAGGAGnCA 60

55

	TAACAAATAA AGGTGCGTTA TTAATAACAG TGCCAGGCAA AAATGATGAA GTACAACGCT	180
	GTATTACTGC TCATGTTGAT ACTTTAGGTG CaATGGTTAA AGAAATTAAA GAAGATGGTC	240
5	GCTTaGCAAT AGAATTAATT GGAGGATTCA CGTATAACGC GATTGAGGGT GAATATTGCC	300
	AAATTAAAAC TGATGCTGGT CAAATATATA CAGGAACAAT TTGTCTGCAT GAAACAAGTG	360
10	TTCATGTATA TAGAAATAAT CATGAAATAC CTAGAGATCA AAAGCATATG GAAATAAGAA	420
	TTGATGAAGT AACTACATCA GAAGAAGATA CAAAGAGTTT AGGTATTTCA GTAGGTGATT	480
	TTGTTAGCTT TGATCCACGT ACAGTTATCA CGTCATCAGG TTTTATTAAA TCTCGTCATT	540
15	TAGATGATAA AGCTAGCGTA CGgTtGATAC TACAATTACT AAAGAAATTA AAAGAAGAGC	600
	AAATAATATT ACCACATACA ACGCAATTTT ATATTTCTAA TAACGAAGAA ATAGGTTACG	660
	GTGCAAATGC ATCAATTGAT TCGAAAATCA AAGAATATAT TGCATTAGAT ATGGGCGCGT	720
20	TGGGAGACGG TCAAGCATCG GATGAATATA CAGTTTCTAT TTGTGCCAAA GATGCTTCAG	780
	GTCCATATCA TAAGCAATTG AAATCGCACC TAGTTAATCT TTGCAAAATA AATAACATTC	840
	CATATAAAGT AGACATATAT CCATATTATG GTTCAGATGC TTCAGCAGCT TTACATGCTG	900
25	GTGCGGATAT CAGACATGGT TTATTTGGCG CTGGCATTGA ATCATCTCAT GCAATGGAAC	960
	GAACACATAT TGATTCTATT AAAGCGACAG AGAAATTACT ATATGCATAT TGCTTATCAC	1020
30	CAATTGAGTA AACAATTAGT GTTGACAAAT GTGaACGACC TATGTAATAT AATGAACTAT	1080
	AAAAATAATT AGAATTTTCT AAAGAAATAG TAGCAGATAT GAAACGTAGC AAATAGAAAG	1140
	CTAATGGGTG ATGGGAATTA GCACGCCATA TCTTGTAAT TGGACTTTGG AAAACAATTG	1200
35	AATGAGTTTT GAAAGTGAAC ATGAATTATG TTAAC TAAGG TGGCACCACG GTAACGCGTC	1260
	CTTACAGGTA TATGCGTTAT GTGGTGTCTT TTTATTTAGA CAAAATGTAG TAGTTAATTA	1320
	AAGGTAGCAA CAGAAAGTTA GTGGATGATG TGAAC TAACA CCGAGATTAA TGAAATTGGG	1380
40	TTTTGTCTGC AACAGAAAAA TTATATATAG TAAAGAGTGA ACTATGAATA TTTCGAATAT	1440
	TCGGTTAATT TAGGTGGTAC CACGCGTCAC nTCCTTTATA TTGATAAGGA TGCTGGCGCT	1500
	TTTTTGAAAG GAGCGTATAG AATGGATATA TTTTATAAAA AAATAAAAGC AAATGTAACG	1560
45	CCCGAAGTTT TAGCACAACT TCATTCCAAG AAGaTCATTT TGGAAAGTAC AAATCAACAA	1620
	CAAACTAAAG GTCGCTATTC AGTTGTTATT TTTGATATTT ATGGCACTTT AACTTTAGAT	1680
50	AATGATGTAT TATCAGTAAG TACTTTAAAA GAATCGTATC AAATCACTGA AAGACCGTAC	1740
	CATTATTTAA CGACTAAaAT AAATGAAGAC TACCATAATA TTCCAAGATG AGGCAACTTA	1800
	AGTCATTA	1808
55		

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1320 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

10 TGGTCGTCAA TTTCTTGATT ATATCTATAA TCCTCATTTT CAATATTAGA GTCTGTAGAA 60  
 TCATCGATAT TATTATCATT CGCATGACTA GAAGCAGAAT CATTATTTTT ATCATTGCTT 120  
 15 TCTTCTTTTT TGAAGTCTTT ATTTATCAAG TAAATTTCTT CATCAAAATC AGCTTGTTGA 180  
 GATGTATCAT CTTTATTTTG ATTAGAAAAA TGTGTAGCCT TTGATCTTTT TCTTTGCCGT 240  
 CTTTCTTAG ATGTATTCCT CGTAAATAAT TCTAATTCAT CTTTATCTTC ATTGATTCT 300  
 20 TGTGTATCGT TCTTCGTTTT ATCATCCATC AATACTCACA CCCTTTAATA AGATGGTAA 360  
 TGGGCACGGA ATCTTTCAAT AAATTTCTCT CCACGCTCTT CAAAAGTACT ATATTGATCC 420  
 CAACTCGCAC AAGCAGGTGA CAATAATACA ACATCATTTG GTTCTATAAT ATCTTGTA 480  
 25 TTATCAACAG CGTCTTCGAC ATTGTTGCT TCAATGACCG ATTCCCTTG ACTATTACCT 540  
 AGTTTAGCAA ACTTAGCTTT CGTTTGTCGG AATACAACCA TCGCGCGAAC ATTTCCATA 600  
 TAAGGAATGA GTTCGTCAA TTCATTCCCT CGATCCAAAC CACCACATAA CCAAATGATT 660  
 30 GGTGATTAA ATGAATTTAA GGCAAAGTGT GTTGCTAGCG TGTTTGTTGC TTTGGAATCA 720  
 TTATAATATT TATTAGTTCT ATTAGTACCA ACATATTGCA ATCTATGCTC TATTCCTGAA 780  
 35 AATGTAGTTA AACTATCAAT AATTGCTTTA ATAGGTACAC CAGCANAATA CAAGCAAGCA 840  
 CAGCTGCTAA TATATTTCTA AATTATGTTT ACCAGGCAAT ACTAGATCTT CAGTGTTAAT 900  
 AATACTGAACA CCTTTATAAA CGATAAAACC ATCTTTAATA TAAATACCAT CATCTCTTG 960  
 40 TTGAGTTGAG AAATACAATG TCTTAGCTTT TAATCTTCC GACTCTATCA CTTGTCTTTG 1020  
 ATGATAATTA CAAATCAAAT AATCCTCTTC CGTTTGATTT TTATATATTT GCTTTTLAGC 1080  
 ATTTTGATAG TTTTCTAAAT TTTCATGGTA ATCTAGATGC GCCGAATAAA TGTTAGTAAT 1140  
 45 TATAGCAATG TGTGGTTTAT ACTTTTCGAT TCCAAGTAAC TGAATGACG ACAACTCTGT 1200  
 AACTAAATAA TCTGTAGGCT TTACTTCTTG TGCTACTTTA GATGCAACAT AACCAATATT 1260  
 50 GCCGGATAAT CTTCCAGTTA AGCGACTTTT TTTAAACATA TCTCCAATTA GAGAAGTAAC 1320

## (2) INFORMATION FOR SEQ ID NO: 81:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4280 base pairs

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

	TTTACACCAA TCAAAAAATC GAACTGATAT AAATAAGTAC AAAGCTTATC TATCAATCCG	60
10	ATTTAGTTAT AAAACAAAAA AAGCCACAGT AATGTGGCTT TTTGTTATAT TCAGTATCAA	120
	AATGGTATCA ATAGCCATTT TCGGAAGTCA AGAATGGCTT AACAAACGCGG TTTAAAGCTA	180
	TCCAATACTA CCTTCCATTT CGAACTTGAT TAAACGGTTC ATTTGACCG CGTATTCCAT	240
15	TGGAAGTTCT TTTGTAAATG GTTCGATGAA TCCCATAACA ATCATTCTCTG TCGCTTCTTC	300
	TTCAGAAATA CCACGACTCA TTAGATAGAA TAATTGTTCT TCAGAAACTT TTGAAACCTT	360
	GGCTTCATGT TCTAATGATA TTTGATCGTT GAATACTTCG TTATATGGAA TTGTATCTGA	420
20	TGTTGATTCTG TTATCTAAGA TTAATGTATC ACATTCAATA TTTGAACGAG CACCTTTTGC	480
	TTTACGTCCA AAATGAACAA TACCGCGATA AATAACTTTA CCACCATTTT TAGAAATAGA	540
	TTTAGAAACA ATTGTAGAAG ATGTATTAGG TGCTTTATGA ATCATTTTAG CACCGGCATC	600
25	TTGAACTTGT CCTTTACCAG CAAATGCAAT AGATAATGTA CTACCTTTTG CACCTTCACC	660
	TAAAAGAACA CAGTTTGGAT ATTTTCATCGT TAACTTAGAA CCTAAGTTAC CATCTACCCA	720
30	TTCCATATTT CCGTTTTTCAT AAACAAAAGT ACGTTTTGTA ACTAAATTGT ATACATTGTT	780
	CGCCCAGTTT TGAATCGTAG TATAACGAAC GTGCGCATCT TTATGCACAA TGATTTCAC	840
	AACAGCAGAG TGTAAGAAG TAGTTGTATA AACTGGTGCA GTACAACCTT CTACGTAATG	900
35	TACAGAAGCA CCTTCATCAG CAATGATTAA TGTACGTTCA AATTGACCCA TGTCTCAGA	960
	GTTAATACGG AAATAAGCTT GTAGTGGCGT ATCTAGTTTG ATATTTTATG GTACATAAAT	1020
	GAAGGAACCA CCGTACCATA CTGCTGAGTT TAACGCCGCA AATTTGTTAT CTGCTGCAGG	1080
40	TACTACAGAA GCAAAGTATT TTTTGAATAA TTCTTCATTT TCTTGTAAG CACTATCTGT	1140
	ATCTTTAAAG ATAATACCTT TTTCTTCAAG TTCTTTTCC ATATTATGGT AAACAACCTC	1200
	AGATTCATAT TGAGCAGAAA CACCAGCTAA ATATTTTGTG TCAGCTTCAG GAATTCCTAA	1260
45	TTTATCGAAA GTTCTTTTAA TTTCTTCTGG CACTTCATCC CATGAACGTT CAGCTTGTTT	1320
	TGAAGGCTTT ACATAGTAAG TAATGTCATC GAAATTCAT TCTGATAAGT CGCCACCCCA	1380
50	TTGAGGCATT GGCATTTTAT AAAACAATTT TAATGATTTA AGACGGAAAT CTAACATCCA	1440
	TTCCGGCTCA TTTTTCATGT TAGAAATTTT TCTAACGATA TTCTCAGTTA AACCACGTTT	1500
	TGATCTGAAA ATGGACACAT CATCGTCGTG GAATCCATAT TTATAATCCC CAACATCAGG	1560

55



	TTTAATTCAT GATGTAAACC ATATTATAAC AATGACATGA CATCTTATAA AAATTTTAT	1680
	ACTTTTATAT GTCTAATATC AAAATTATCT ATGATTAACA GCATTCTATT CTTCTTCAGT	1740
5	CGTACCTTCT GCTTTACCTT CTTTAGCAAC AGTACCTTTT TCCAATGCTT TCCAAGCTAA	1800
	TGTGGCACAT TTAATACGAG CTGGGAATTG AGATACACCT TGCAATGCTT CAATATCTCC	1860
	CATTCTTCT GTAATCACAT AGTCTTCACC AAGCATCATT TTCGTAAATT CTGGGCTCAT	1920
10	TTGCATTGCT TCTCCAAGTG AATGACCTTT AACAGCTTGT GTCATCATCG ATGCACTTGC	1980
	CATTGAAATC GAACAACCTT CACCTTCAAA CTTAGCATCT TTTATAATGC CGTCTTCTAT	2040
15	ATCAAATGTT AGTCGTATAC GGTCAACGCA TGTGCGGTTA TTCATATCTA CTGTCATAGA	2100
	CCCGTTATCT AATACACCTT TATTTCTAGG ATTTTATAA TGATCCATAA TGACAGATCT	2160
	ATATAATTGA TCTAGATTAT TAAAATTCAT AAGAGAAAAA CTCCTTCGTT TGTTC AAGG	2220
20	CATTATTAA CTGATCAACG TCTTCTTCG TGTGTATAT ATAAAACTC GCTCTAGCTG	2280
	TTGAAGACAC ATTTAACCAT TTCATTAACG GTTGCGCACA ATGATGCCCA GCTCTAACCG	2340
	CTACACCTTC TGTATCTACG GCTGTAGCAA CATCGTGTGG ATGTACATCT TGTAAATTAA	2400
25	ACGTTATTAC ACCTGCACGA CGATCCTTTG GCGGGCCATA AATTCAATT CCTTCAATTG	2460
	CAGACATTTG CTCATAAGCA TATATCGTTA ATTCTTGTTC ATATTATGA ATTGCATCAA	2520
	AACCTATGCG TTCTAAATAG CGAATAGCTT CTGCAAGCCC AATTGCTTGA GCAATTAATG	2580
30	GAGTACCCGC CTCAAATTTA GTAGGTAAAT CAGCCCATGT TGCATCATAC TTACTTACAA	2640
	AATCAATCAT GTCGCCACCG AACTCAATCG GTTCCATTTT TTGTAGTAAC TCACGTTTAC	2700
	CAAATAATAC GCCAATACCT GTTGGTCCAA GCATTTTATG ACCACTAAAA CTATAAAAT	2760
35	CAGCATT CAT TTCTTG CATA TCAAGTTTCA TATGTGGTGC TGCTTGCGCC CCATCAACAC	2820
	TGATTAATTGC ACCATGTTGA TGAGCTATTT CTGCAATGGT TTTAACATCA TTAATTGTAC	2880
40	CGAGCACATT AGATATATGT GCAATAGCAA CGATCTTTGT TTTATCATTA ATCGTTTGCT	2940
	TAATATCCTC GATGTTTAAAT TCACCGTCAG CTGTCATTGG TATAAATTC AATGTCGCAT	3000
	TTTTACGCTT TGCTAACTGT TGCCAAGGAA CAATATTGGC ATGATGTTCC ATTTCACTGA	3060
45	CAACAATTTT ATCGCCCTCT TCAACATTTG CATCACCATA GCTATGTGCT ACAAGGTTAA	3120
	TCGACGCAGT TGTTCCGCGT GTAAAAATGA TTTCTTCAAA ATACTTCGCA TTAATAAAAC	3180
	GACGAACGGT TTCACGGGCA TTTTCATAAC CATCAGTTGC CAATGATCCT AATGTATGAA	3240
50	CACCACGATG AACGTTTGAA TTATAACGCT TGTAGTAATC TTCTAAAACA TTTAACACTT	3300
	GCACAGGCGT TTGACTTGTC GCTGTTGAAT CAAGATATGC TAAACGTTTG CCATTGACTT	3360
55		

CTTCATTCAC GACCTTTCTT AAATAAAAAAT CCTAATCATT TAAATACTGA CGTTGTATTA 3480  
 GTCTTATACC AATATCGACA GTCTATATCT ATTACAAACT TTTATTTTCA AAATATTATT 3540  
 5 TAGAAACTTT GCGTTCAATT ACTTCTCTCA ATTGACGTTT AACGTCTTCG ATAGGTAATT 3600  
 CACGTACTAC TGGATCTAAG AAACCATGTA TAACAAGACG TTCCGCTTCT CTTTGAGAAA 3660  
 TACCACGACT CATTAAATAG TAAAGTTGAT CTGGATCAAC ACGACCTACT GATGCAGCAT 3720  
 10 GACCAGCTTG TACATCATCT TCATCAATTA ATAAAATAGG ATTCGCGTCA CCACGAGCAT 3780  
 GTTCAGATAA CATTAAATACA CGTGATTCTT GATTAGCAAT TGATTTAGTT CCACCATGCT 3840  
 TAATGTAGCC GATACCATTA AATACAGACG ATGCATGTTT TTTATAACA CCATGTTTAA 3900  
 15 GGATATAACC ATCTGTTTCT TTACCATATT GTACGATTTT AGATGTTAGA TTAATTTTTT 3960  
 GTTCGCCTGT ACCTACAAC ACTGATTTAA GTGAACCTGT TGAACGATCA CCAAATAAAT 4020  
 20 TTGTTGTATT ATCAATAATT TGGCTACCCT CATTCATTAA ACCTAGTGCC CAATTAATTG 4080  
 AGGCATCCGC TTCAGTAATA CCACGTCGAA TGATATGACC TGTAAGCCT TTATCCATAT 4140  
 AGTCCACTGA GCCATATGTG ATATTGAAT TTGCACCAGC AATCACTTCA GAAATAATAT 4200  
 25 TCAATTGATT TCCTTCACCA GATGCATTTG mTAAGTAATT TTCAACATAT GTGACTTCGG 4260  
 CGCTTTCTTC AGTAACGATG 4280

## (2) INFORMATION FOR SEQ ID NO: 82:

- 30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15598 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

40 TChGACTCGA ACGGTGmAAC TAttCCGTTG TaATTCCgGA GgAAaCAAGG TATGCCCATC 60  
 TGCaaAGAAA gaATGsAATG AACTTTTTGG AAATGTAGAA GTGGTAAATA AAGATAAAGG 120  
 ATATTACATT CTGAGAAGTA TAAAAGCTTG AAATGAAATG GATATTCTGT TATAGTTATA 180  
 45 TAATGTAAAA ATTTATGTTT AATAAGTGTG TACTTTTACG TTAAATAGAT AAGTTAATTA 240  
 AGAATAAATA TAGAATCGAA AATGGTGTCA TCATTAGTGT TGCCGTTTTT TTTTGTCTT 300  
 TTTATTAATA TGCTTATGGT ATTTAGCTAA AAGCGGATCA CATAATTTTT GAGGGGTGAA 360  
 50 TCTGTTTGGC AGGTCAAGTT GTCCAATATG GAAGACATCG TAAACGTAGA AACTACGCGA 420  
 GAATTTTACA AGTATTAGAA TTACCAAACT TAATAGAAAT TCAAATAA TCTTACGAGT 480

55

EP 0 786 519 A2

	CTGGTAATTT	GTCATTAGAG	TTTGTGGATT	ACCGTTTAGG	AGAACCAAAA	TATGATTTAG	600
	AAGAATCTAA	AAACCGTGAC	GCTACTTATG	CTGCACCTCT	TCGTGTAAAA	GTGCGTCTAA	660
5	TCATTAAAGA	AACAGGAGAA	GTTAAAGAAC	AAGAAGTCTT	TATGGGTGAT	TTCCCATTAA	720
	TGACTGATAC	AGGTACGTTT	GTTATCAATG	GTGCAGAACG	TGTAATCGTA	TCTCAATTAG	780
	TTGCTTCACC	ATCCGTTTAT	TTCAATGAAA	AAATCGACAA	AAATGGTCGT	GAAAACTATG	840
10	ATGCAACAAT	TATTCCAAAC	CGTGGTGTCAT	GGTTAGAATA	TGAAACAGAT	GCTAAAGATG	900
	TTGTATACGT	ACGTATTGAT	AGAACACGTA	AACTACCATT	AACAGTATTG	TTACGTGCAT	960
15	TAGGTTTCTC	AAGCGACCAA	GAAATTGTTG	ACCTTTTAGG	TGACAATGAA	TATTTACGTA	1020
	ATACTTTAGA	GAAAGACGGC	ACTGAAAACA	CTGAACAAGC	GTTATTAGAA	ATCTATGAAC	1080
	GTTTACGTCC	AGGTGAACCA	CCAACGTGTT	AAAATGCTAA	AAGTCTATTG	TATTCACGTT	1140
20	TCTTTGATCC	AAAACGCTAT	GACTTAGCAA	GCGTGGGTCG	TTATAAAACA	AACAAAAAAT	1200
	TACATTTAAA	ACATCGTTTA	TTTAATCAAA	AATTAGCTGA	GCCAATTGTA	AATACTGAAA	1260
	CTGGTGAAAT	TGTAGTTGAA	GAAGGTACAG	TGCTTGATCG	TCGTAAATC	GACGAAATCA	1320
25	TGGATGTACT	TGAATCAAAT	GCAAACAGCG	AAGTGTGTTG	ATTGCATGGT	AGCGTTATAG	1380
	ACGAGCCAGT	AGAAATTCAA	TCAATTAAAG	TATATGTTCC	TAACGATGAT	GAAGGTGCTA	1440
	CGACAACGT	AATTGGTAAT	GCTTTCCCTG	ACTCAGAAGT	TAAATGCATT	ACACCAGCAG	1500
30	ATATCATTCG	TTCAATGAGT	TACTTCTTTA	ACTTATTAAG	CGGTATTGGA	TATACAGATG	1560
	ATATTGACCA	TTTAGGTAAC	CGTCGTTTAC	GTTCTGTAGG	TGAATTACTA	CAAAACCAAT	1620
	TCCGTATCGG	TTTATCAAGA	ATGGAAAGAG	TTGTACGTGA	AAGAATGTCA	ATTCAAGATA	1680
35	CTGAGTCTAT	CACACCTCAA	CAATTAATTA	ATATTGACCC	TGTTATTGCA	TCTATTAAAG	1740
	AATTCTTTGG	TAGCTCTCAA	TTATCACAAT	TCATGGACCA	AGCAAACCCA	TTAGCTGAGT	1800
40	TAAAGCATAA	ACGTCGTCTA	TCAGCATTAG	GACCTGGTGG	TTTAACACGT	GAACGTGCTC	1860
	AAATGGAAGT	ACGTGACGTT	CACTACTCTC	ACTATGGCCG	TATGTGTCCA	ATTGAAACAC	1920
	CTGAGGGACC	AAACATTGGA	TTGATTAACT	CATTATCAAG	TTATGCACGT	GTAAATGAAT	1980
45	TCGGCTTTAT	TGAAACACCA	TATCGTAAAG	TTGATTTAGA	TACACATGCT	ATCACTGATC	2040
	AAATTGACTA	TTTAACAGCT	GACGAAGAAG	ATAGCTATGT	TGTAGCACAA	GCAAACTCTA	2100
	AATTAGATGA	AAATGGTCGT	TTCATGGATG	ATGAAGTTGT	ATGTCGTTTC	CGTGGTAACA	2160
50	ATACAGTTAT	GGCTAAAGAA	AAAATGGATT	ATATGGATGT	ATCGCCGAAG	CAAGTTGTTT	2220
	CAGCAGCGAC	AgcATGTATT	CCATTCTTAG	AAAATGATGA	CTCAAACCGT	GCATTGATGG	2280
55							

	CAGGTATGGA ACACGTTGCA GCACGTGATT CTGGTGCGGC TATTACAGCT AAGCACAGAG	2400
	GTCGTGTTGA ACATGTTGAA TCTAATGAAA TTCTTGTTTCG TCGTCTAGTT GAAGAGAACG	2460
5	GCGTTGAGCA TGAAGGTGAA TTAGATCGCT ATCCATTAGC TAAATTTAAA CGTTCAAACT	2520
	CAGGTACATG TTACAACCAA CGTCCAATCG TTGCAGTTGG AGATGTTGTT GAGTATAACG	2580
10	AGATTTTAGC AGATGGACCA TCTATGGAAT TAGGAGAAAT GGCATTAGGT AGAAACGTAG	2640
	TAGTTGGTTT CATGACTTGG GACGGTTACA ACTATGAGGA TGCCGTTATC ATGAGTGAAA	2700
	GACTTGTAAG AGATGACGTG TATACTTCTA TTCATATTGA AGAGTATGAA TCAGAAGCAC	2760
15	GTGATACTAA GTTAGGACCT GAAGAAATCA CAAGAGATAT TCCTAATGTT TCTGAAAGTG	2820
	CACTTAAGAA CTTAGACGAT CGTGGTATCG TTTATATTGG TGCAGAAGTA AAAGATGGAG	2880
	ATATTTTAGT TGGTAAAGTA ACGCCTAAAG GTGTAAGTGA GTTAACTGCC GAAGAAAGAT	2940
20	TGTTACATGC AATCTTTGGT GAAAAAGCAC GTGAAGTTAG AGATACTTCA TTACGTGTAC	3000
	CTCACGGCGC TGGCGGTATC GTTCTTGATG TAAAAGTATT CAATCGTGAA GAAGGCGACG	3060
	ATACATTATC ACCTGGTGTA AACCAATTAG TACGTGTATA TATCGTTCAA AAACGTAAAA	3120
25	TTCATGTTGG TGATAAGATG TGTGGTCGAC ATGGTAACAA AGGTGTCATT TCTAAGATTG	3180
	TTCTGAAGA AGATATGCCT TACTTACCAG ATGGACGTCC GATCGATATC ATGTTAAATC	3240
	CTCTTGGTGT ACCATCTCGT ATGAACATCG GACAAGTATT AGAGCTACAC TTAGGTATGG	3300
30	CTGCTAAAAA TCTTGGTATT CACGTTGCAT CACCAGTATT TGACGGTGCA AACGATGACG	3360
	ATGTATGGTC AACAAATGAA GAAGCTGGTA TGGCTCGTGA TGGTAAAACT GTACTTTATG	3420
35	ATGGACGTAC AGGTGAACCA TTCGATAACC GTATTTTCACT AGGTGTAATG TACATGTTGA	3480
	AACCTGCGCA CATGGTTGAT GATAAATTAC ATGCGCGTTC AACAGGACCA TATTCATTG	3540
	TTACACAACA ACCACTTGGC GGTAAAGCGC AATTCGGTGG ACAACGTTTT GGTGAGATGG	3600
40	AGGTATGGGC ACTTGAAGCA TATGGTGCTG CATACACATT ACAAGAAATC TTAAGTTACA	3660
	AATCCGATGA TACAGTAGGA CGTGTGAAAA CATACGAGGC TATTGTTAAA GGTGAAAAA	3720
	TCTCTAGACC AAGTGTCCA GAATCATTCC GAGTATTGAT GAAAGAATTA CAAAGTTTAG	3780
45	GTTTAGATGT AAAAGTTATG GATGAGCAAG ATAATGAAAT CGAAATGACA GACGTTGATG	3840
	ACGATGATGT TGTAGAACGC AAAGTAGATT TACAACAAAA TGATGCTCCT GAAACACAAA	3900
	AAGAAGTTAC TGATTAATAC GCAATTTACA AAACAGGCAA AAAGATACTA AGCTGAATTT	3960
50	TATTGATGAT TCAGTTTAGT ACTTTAAGCC ATTTTAAATA AATGCAAATC AATCAAATAG	4020
	CACAGCTAAT CTAAATTGAA GGAGGTAGGC TCCTTGATTG ATGTAAATAA TTTCCATTAT	4080

55

	AAACCTGAAA CAATCAACTA CCGTACATTA AAACCTGAAA AAGATGGTCT ATTCTGTGAA	4200
	AGAATTTTCG GACCTACAAA AGACTGGGAA TG TAGTTGTG GTAAATACAA ACGTGTTTCG	4260
5	TACAAAGGCA TGGTCTGTGA CAGATGTGGA GTTGAAGTAA CTAAATCTAA AGTACGTCGT	4320
	GAAAGAATGG GTCACATTGA ACTTGCTGCT CCAGTTTCTC ACATTGGTA TTTCAAAGGT	4380
	ATACCAAGTC GTATGGGATT ATTACTTGAC ATGTCACCAA GAGCATTAGA AGAAGTTATT	4440
10	TACTTTGCTT CTTATGTTGT TG TAGATCCA GGTCCAAC TG GTTTAGAAAA GAAAACTTTA	4500
	TTATCTGAAG CTGAATTCAG AGATTATTAT GATAAATACC CAGGTCAATT CGTTGCAAAA	4560
	ATGGGTGCAG AAGGTATTAA AGATTTACTT GAAGAGATTG ATCTTGACGA AGAACTTAA	4620
15	TTGTTACGCG ATGAGTTGGA ATCAGCTACT GGTCAAAGAC TTACTCGTGC AATTAAACGT	4680
	TTAGAAGTTG TTGAATCATT CCGTAATTCA GGTAAACAAAC CTTCATGGAT GATTTTAGAT	4740
20	GTACTTCCAA TCATCCCACC AGAAATTCGT CCAATGGTTC AATTAGATGG TGGACGATTT	4800
	GCAACAAGTG ACTTAAACGA CTTATACCGT CGTGTAATTA ATCGAAATAA TCGTTTGAAA	4860
	CGTTTATTAG ATTTAGGTGC ACCTGGTATC ATCGTTCAAA ACGAAAAACG TATGTTACAA	4920
25	GAAGCCGTTG ACGCTTTAAT TGATAATGGT CGTCGTGGTC GTCCAGTTAC TGGCCCAGGT	4980
	AACCGTCCAT TAAAATCTTT ATCTCATATG TTAAGGTA AACAAGGTCG TTTCCGTCAA	5040
	AACTTACTTG GTAAACGTGT TGA CTATTCA GGACGTTGAG TTATTGCAGT AGGTCCAAGC	5100
30	TTGAAAATGT ACCAATGTGG TTTACCAAAA GAAATGGCAC TTGAACTATT TAAACCATT	5160
	GTAATGAAAG AATTAGTTCA ACGTGAAATT GCAACTAACA TTA AAAATGC GAAGAGTAAA	5220
	ATCGAACGTA TGGATGATGA AGTTTGGGAC GTATTGGAAG AAGTAATTAG AGAACATCCT	5280
35	GTATTACTTA ACCGTGCACC AACACTTCAT AGACTTGGTA TTCAAGCATT TGAACCAACT	5340
	TTAGTTGAAG GTCGTGCGAT TCGTCTACAT CCACTTGTA CAACAGCTTA TAACGCTGAC	5400
40	TTTGACGGTG ACCAAATGGC GGTTCACGTT CCTTTATCAA AAGAGGCACA AGCTGAAGCA	5460
	AGAATGTTGA TGTTAGCAGC ACAAACATC TTGAACCCTA AAGATGGTAA ACCTGTAGTT	5520
	ACACCATCAC AAGATATGGT ACTTGGTAA TATTACCTTA CTTTAGAAAG AAAAGATGCA	5580
45	GTAAATACAG GCGCAATCTT TAATAATACA AATGAAGTAT TAAAAGCATA TGCAAATGGC	5640
	TTTGTACATT TACACACTAG AATTGGTGTA CATGCAAGTT CGTTCAATAA TCCAACATTT	5700
	ACTGAAGAAC AAAACAAAAA GATTCTTGCT ACGTCAGTAG GTAAAATTAT ATTCAATGAA	5760
50	ATCATTCCAG ATTCATTTGC TTATATTAAT GAACCTACGC AAGAAAACTT AGAAAGAAAG	5820
	ACACCAAACA GATATTTTCAT CGATCCTACA ACTTTAGGTG AAGGTGGATT AAAAGAATAC	5880
55		

	GAAGTATTCA ACAGATTAG CATCACTGAT ACATCAATGA TGTTAGACCG TATGAAAGAC	6000
	TTAGGATTCA AATTCTCATC TAAAGCTGGT ATTACAGTAG GTGTTGCTGA TATCGTAGTA	6060
5	TTACCTGATA AGCAACAAAT ACTTGATGAG CATGAAAAAT TAGTCGACAG AATTACAAAA	6120
	CAATTCAACC GTGGTTTAAT CACTGAAGAA GAAAGATATA ATGCAGTTGT TGAAATTTGG	6180
	ACAGATGCAA AAGATCAAAT TCAAGGTGAA TTGATGCAAT CACTTGATAA AACTAACCCA	6240
10	ATCTTCATGA TGAGTGATTG AGGTGCCCGT GGTAAACGCAT CTAACCTTAC ACAGTTAGCA	6300
	GGTATGCGTG GATTGATGGC CGCACCATCT GGTAAAGATTA TCGAATTACC AATCACATCT	6360
15	TCATTCCGTG AAGGTTTAAC AGTACTTGAA TACTTCATCT CAACTCACGG TGCACGTAAA	6420
	GGTCTTGCCG ATACAGCACT TAAAACAGCT GACTCAGGAT ATCTTACTCG TCGTCTTGTT	6480
	GACGTGGCAC AAGATGTTAT TGTTCGTGAA GAAGACTGTG GTACTGATAG AGGTTTATTA	6540
20	GTTTCTGATA TTAAAGAAGG TACAGAAATG ATTGAACCAT TTATCGAACG TATTGAAGGT	6600
	CGTTATTCTA AAGAAACAAT TCGTCATCCT GAAACTGATG AAATAATCAT TCGTCTGAT	6660
	GAATTAATTA CACCTGAAAT TGCTAAGAAA ATTACAGATG CTGGTATTGA ACAAATGTAT	6720
25	ATTCGCTCAG CATTTACTTG TAACGCACGA CATGGTGTTC GTGAAAAATG TTACGGTAAA	6780
	AACCTTGCTA CTGGTGAAAA AGTTGAAGTT GGTGAAGCAG TTGGTACAAT TGCAGCCCAA	6840
	TCTATCGGTG AACCAGGTAC ACAGCTTACA ATGCGTACAT TCCATACAGG TGGGGTAGCA	6900
30	GGTAGCGATA TCACACAAGG TCTTCCTCGT ATTCAAGAGA TTTTCGAAGC ACGTAACCcT	6960
	AAAGGTCAAG CGGTAATTAC GGAAATCGAA GGTGTCGTAG AAGATATTAA ATTAGCAAAA	7020
	GATAGACAAC AAGAAATTGT TGTTAAAGGT GCTAATGAAA CAAGATCATA CCTTGCTTCA	7080
35	GGTACTTCAA GAATTATTGT AGAAATCGGT CAACCAAGTTC AACGTGGTGA AGTATTAACT	7140
	GAAGGTTCTA TTGAACCTAA GAATTACTTA TCTGTGCTG GATTAAACGC GACTGAAAGC	7200
	TACTTATTAA AAGAAGTACA AAAAGTTTAC CGTATGCAAG GTGTAGAAAT CGACGATAAA	7260
40	CACGTTGAGG TTATGGTTCG ACAAATGTTA CGTAAAGTTA GAATTATCGA AGCAGGTGAT	7320
	ACGAAGTTAT TACCAGGTTT ATTAGTTGAT ATTCATAACT TTACAGATGC AAATAGAGAA	7380
45	GCATTTAAAC ACCGTAAGCG TCCTGCAACA GCTAAACCAG TATTACTTGG TATTACTAAA	7440
	GCATCACTTG AAACAGAAAG TTTCTTATCT GCAGCATCAT TCCAAGAAAC AACAGAGTT	7500
	CTTACAGATG CAGCAATTAA AGGTAAGCGT GATGACTTAT TAGGTCTTAA AGAAAACGTA	7560
50	ATTATTGGTA AGTTAATTCC AGCTGGTACT GGTATGAGAC GTTATAGCGA CGTAAAATAC	7620
	GAAAAACAG CTAAACCAGT TGCAGAAGTT GAATCTCAAA CTGAAGTAAC GGAATAACAA	7680
55		

EP 0 786 519 A2

	ATGTTGACGA ATTCTCTTGT TCAATGTAA TATATTAAAG GTTGATGCAA GCAGAACTTT	7800
	GGAGGATAAA TTATTGTCTA AGGAAAAAGT tGCACGCTTT AACAAACAAC ATTTTGTAGT	7860
5	TGGTCTTAAA GAAACGCTTA AAGCGTTAAA GAAAGATCAA GTTACATCTT TGATTATTGC	7920
	TGAAGACGTT GAAGTATATT TAATGACTCG CGTGTTAAGC CAAATCAATC AGAAAAATAT	7980
10	ACCTGTATCT TTTTTCAAAA GCAAACATGC TTTGGGTAAA CATGTAGGTA TTAACGTCAA	8040
	TGCGACAATA GTAGCATTGA TTAAATGAGA ATTAGTAAGT GTTTTACTTA CTAAATTTTA	8100
	TTTAACCTAA AAATGAACCA CCTGGATGTG TGGGATTAAA AAGTGAAGAG AGGAGGACAT	8160
15	ATCACATGCC AACTATTAAC CAATTAGTAC GTAAACCAAG ACAAAGCAAA ATCAAAAAAT	8220
	CAGATTCTCC AGCTTTAAAT AAAGGTTTCA ACAGTAAAAA GAAAAAATTT ACTGACTTAA	8280
	ACTCACCACA AAAACGTGGT GTATGTACTC GTGTAGGTAC AATGACACCT AAAAAACCTA	8340
20	ACTCAGCGTT ACGTAAATAT GCACGTGTGc gTtTATCAAA CAACATCGAA ATTAACGCAT	8400
	ACATCCCTGG TATCGGACAT AACTTACAAG AACACAGTGT TGTACTTGTA CGTGGTGGAC	8460
	GTGTAAAAGA CTTACCAGGT GTGCGTTACC ATATTGTACG TGGAGCACTT GATACTTCAG	8520
25	GTGTTGACGG ACGTAGACAA GGTGTTTCAT TATACGGAAC TAAGAAACCT AAAAACTAAG	8580
	AATTTAGTTT TTAATTAAAT CTAAACTTA AAATATTTAA TATAAGGAAG GGAGGATTTA	8640
	CATTATGCCT CGTAAAGGAT CAGTACCTAA AAGAGACGTA TTACCAGATC CAATTCATAA	8700
30	CTCTAAGTTA GTAATAAAT TAATTAACAA AATTATGTTA GATGGTAAAC GTGGAACAGC	8760
	ACAAAGAATT CTTTATTTCAG CATTTCGACCT AGTTGAACAA CGCAGgtTCG TGATGCATTA	8820
	GAAGTATTCG AAGAAGCAAT CAACAACATT ATGCCAGTAT TAGAAGTTAA AGCTCGTCGC	8880
35	GTAGGTGGTT CTAACATCA AGTACCAGTA GAAGTTCGTC CAGAGCGTCG TACTACTTTA	8940
	GGTTTACGTT GGTTAGTTAA CTATGCACGT CTTCTGTTGTG AAAAAACGAT GGAAGATCGT	9000
40	TTAGCTAACG AAATTTTAGA TGCAGCAAAT AATACAGGTG GTGCCGTTAA GAAACGTGAG	9060
	GACACTCACA AAATGGCTGA AGCAAACAAA GCATTTGCTC ACTACCGTTG GTAAGATAAA	9120
	AGCTTTTACC CTGAGTGTGT TCTATATTAA TGAATTTTCA TTAAGCGTTC ATGCTTAGGG	9180
45	CATCGCCATA TCTATCGTAT TTATTCAGTA ATATAAACTG GAAGGAGAAA AAATACATGG	9240
	CTAGAGAATT TTCATTAGAA AAAACTCGTA ATATCGGTAT CATGGCTCAC ATTGATGCTG	9300
	GTAAAACGAC TACGACTGAA CGTATTCTTT ATTACACTGG CCGTATCCAC AArGknGGTG	9360
50	AAaCACACGA AGGTGCTTCA CAAATGGACT GGATGGAGCA AGAACAAGAC CGTGGTATTA	9420
	CTATCACATC TGCTGCAACA ACAGCAGCTT GGAAGGTCA CCGTGTAAC ATTATCGATA	9480
55		

	CAGTTACAGT	ACTTGATGCA	CAATCAGGTG	TTGAACCTCA	AACTGAAACA	GTTTGGCGTC	9600
	AGGCTACAAC	TTATGGTGTT	CCACGTATCG	TATTTGTAAA	CAAAATGGAC	AAATTAGGTT	9660
5	CTAACTTCGA	ATACTCTGTA	AGTACATTAC	ATGATCGTTT	ACAAGCTAAC	GCTGCTCCAA	9720
	TCCAATTACC	AATTGGTGCG	GAAGACGAAT	TCGAAGCAAT	CATTGACTTA	GTTGAAATGA	9780
	AATGTTTCAA	ATATACAAAT	GATTTAGGTA	CTGAAATTGA	AGAAATTGAA	ATTCCTGAAG	9840
10	ACCACTTAGA	TAGAGCTGAA	GAAGCTCGTG	CTAGCTTAAT	CGAAGCAGTT	GCAGAAACTA	9900
	GCGACGAATT	AATGGA AAAA	TATCTTGGTG	ACGAAGAAAT	TTCAGTTTCT	GAATTAAAAAG	9960
	AAGCTATCCG	CCAAGCTaCt	AcTAACGTAG	AATTCTACCC	AGTACTTTGT	GGTACAGCTT	10020
15	TCAAAAACAA	AGGTGTTCAA	TTAATGCTTG	ACGCTGTAAT	TGATTACTTA	CCTTCACCAC	10080
	TAGACGTTAA	ACCAATTATT	GGTCACCGTG	CTAGCAACCC	TGAAGAAGAA	GTAATCGCGA	10140
20	AAGCAGACGA	TTAGCTGAA	TTGCTGCGAT	TAGCGTTCAA	AGTTATGACT	GACCCCTTATG	10200
	TTGGTAAATT	AACATTCTTC	CGTGTGTATT	CAGGTACAAT	GACATCTGGT	TCATACGTTA	10260
	AGAACTCTAC	TAAAGGTAAA	CGTGAACGTG	TAGGTCGTTT	ATTACAAATG	CACGCTAACT	10320
25	CACGTCAAGA	AATCGATACT	GTATACTCTG	GAGATATCGC	TGCTGCGGTA	GGTCTTAAAG	10380
	ATACAGGTAC	TGGTGATACT	TTATGTGGTG	AGAAAAATGA	CATTATCTTG	GAATCAATGG	10440
	AATTC CCAGA	GCCAGTTATT	CACTTATCAG	TAGAGCCAAA	ATCTAAAGCT	GACCAAGATA	10500
30	AAATGACTCA	AGCTTTAGTT	AAATTACAAG	AAGAAGACCC	AACATTCCAT	GCACACACTG	10560
	ACGAAGAAAC	TGGACAAGTT	ATCATCGGTG	GTATGGGTGA	GCTTCACTTA	GACATCTTAG	10620
	TAGACCGTAT	GAAGAAAGAA	TTCAACGTTG	AATGTAACGT	AGGTGCTCCA	ATGGTTTCAT	10680
35	ATCGTGAAAC	ATTCAAATCA	TCTGCACAAG	TTCAAGGTAA	ATTCTCTCGT	CAATCTGGTG	10740
	GTCGTGGTCA	ATACGGTGAT	GTTACATTG	AATTCACACC	AAACGAAACA	GGCGCAGGTT	10800
40	TCGAATT CGA	AAACGCTATC	GTTGGTGGTG	TAGTTCCTCG	TGAATACATT	CCATCAGTAG	10860
	AAGCTGGTCT	TAAAGATGCT	ATGGAAAATG	GTGTTTTAGC	AGGTTATCCT	TTAATTGATG	10920
	TTAAAGCTAA	ATTATATGAT	GGTTCATACC	ATGATGTCGA	TTCATCTGAA	ATGGCCTTCA	10980
45	AAATTGCTGC	ATCATTAGCA	CTTAAAGAAG	CTGCTAAAAA	ATGTGATCCT	GTAATCTTAG	11040
	AACCAATGAT	GAAAGTAACT	ATTGAAATGC	CTGAAGAGTA	CATGGGTGAT	ATCATGGGTG	11100
	ACGTAACATC	TCGTCGTGGA	CGTGTGATG	GTATGGAACC	TCGTGGTAAT	GCACAAGTTG	11160
50	TTAATGCTTA	TGTACCACTT	TCAGAAATGT	TCGGTTATGC	AACATCATTA	CGTTCAAACA	11220
	CTCAAGGTCG	CGGTACTTAC	ACTATGTACT	TCGATCACLA	TGCTGAAGTT	CCaAAATCaA	11280

55



EP 0 786 519 A2

	GCCTAGGTTA AAATACAAGG TGAGCTTAAA TGTAAGCTAT CATCTTTATA GTTTGATTTT	11400
	TTGGGGTGAA TGCATTATAA AAGAATTGTA AAATTCTTTT TGCATCGCTA TAAATAATTT	11460
5	CTCATGATGG TGAGAACTA TCATGAGAGA TAAATTTAAA TATTATTTTT AATTAGAATA	11520
	GGAGAGATTT TATAATGGCA AAAGAAAAAT TCGATCGTTC TAAAGAACAT GCCAATATCG	11580
	GTACTATCGG TCACGTTGAC CATGGTAAAA CAACATTAAC AGCAGCAATC GCTACTGTAT	11640
10	TAGCAAAAAA TGGTGACTCA GTTGCACAAT CATATGACAT GATTGACAAC GCTCCAGAAG	11700
	AAAAAGAACG TGGTATCACA ATCAATACTT CTCACATTGA GTACCAAACG GACAAACGTC	11760
	ACTACGCTCA CGTTGACTGC CCAGGACACG CTGACTACGT TAAAAACATG ATCACTGGTG	11820
15	CTGCTCAAAT GGACGGCGGT ATCTTAGTAG TATCTGCTGC TGACGGTCCA ATGCCACAAA	11880
	CTCGTGAACA CATTCTTTTA TCACGTAACG TTGGTGTACC AGCATTAGTA GTATTCTTAA	11940
20	ACAAAGTTGA CATGGTTGAC GATGAAGAAT TATTAGAATT AGTAGAAATG GAAGTTCGTG	12000
	ACTTATTAAG CGAATATGAC TTCCCAGGTG ACGATGTACC TGTAATCGCT GGTTCAGCAT	12060
	TAAAAGCTTT AGAAGGCGAT GCTCAATACG AAGAAAAAAT CTTAGAATTA ATGGAAGCTG	12120
25	TAGATACTTA CATTCCAACG CCAGAACGTG ATTCTGACAA ACCATTTCATG ATGCCAGTTG	12180
	AGGACGTATT CTCAATCACT GGTCTGGTA CTGTTGCTAC AGGCCGTGTT GAACGTGGTC	12240
	AAATCAAAGT TGGTGAAGAA GTTGAAATCA TCGGTTTACA TGACACATCT AAAACAACG	12300
30	TTACAGGTGT TGAAATGTTT CGTAAATTAT TAGACTACGC TGAAGCTGGT GACAACATTG	12360
	GTGCATTATT ACGTGGTGTT GCTCGTGAAG ACGTACAACG TGGTCAAGTA TTAGCTGCTC	12420
	CTGGTTCAAT TACACCACAT ACTGAATTC AAGCAGAAGT ATACGTATTA TCAAAGACG	12480
35	AAGGTGGACG TCACACTCCA TTCTTCTCAA ACTATCGTCC ACAATTCTAT TTCCGTACTA	12540
	CTGAAGTAAC TGGTGTGTT CACTTACCAG AAGGTACTGA AATGGTAATG CCTGGTGATA	12600
	ACGTTGAAAT GACAGTAGAA TTAATCGCTC CAATCGCGAT TGAAGACGGT ACTCGTTTCT	12660
40	CAATCCGTGA AGGTGGACGT ACTGTAGGAT CAGGCGTTGT TACTGAAATC ATTAAATAAT	12720
	TTCTAATTTT TTAGATTTTA TATAAAAAGA AGATCCCTCA ATCGAGGGG CTTTTTTTAA	12780
45	TGTGTAAATT TTGTAATGGC TATTCGATTT AGAAGAACAA TAATTGATGA AAGACTGACT	12840
	AATAAACTT ATAAGTATA ATACTGTTTA AATAAAATTG TTGAGTCTTG GACATTGTAA	12900
	AATGCTCCCT TCAAAGTTTT CATTTTTTCa ATGTCTACTT TGAAGGGAGC ATTTCAATTAG	12960
50	TTTATGTCTC AGATTCATAT CTTTCAATTA ATTTAAATGC TTAATTTGTT TTAAATACTT	13020
	GCTCTAATTC TATGATTTTT AAAAATACAG CTACAGCGTA TTTTAATGAT TTTTCATCAA	13080

55

EP 0 786 519 A2

	TCAGAAAGAA TGCACCTGGT CGTACTTTCA AATAATGTGA AAAATCTTCT CCAATCATCA	13200
	TTAAATCTGA TTCATTAAAG CGTACATGTA AGTCATTGTG TGCTTCTTTA ATAACTTGAT	13260
5	ATGCTTTCTC GTTATTATGG ACAGGCAAAT ACCCTTTAAT ATAATTCAAA TCATAGTTAA	13320
	TATCATTGTC TATTGCTAAA CCTTGTAGAA GCTTATCCAT TTTGTCCATT ACATGATTCT	13380
	GTATATCTGA ATCGAAAGTT CTAAGTGTAC CTTTACAAAA TGCTTGATCA GGAATAACGC	13440
10	TATCTGTGGT GCCTGCTTGA ATCATTCCAA ATGAAAGTAC AGCTTGTTTA ACTGGATCGA	13500
	TCGTACGTGA AATTATTTTT TGTGCACCTA AAATGAACTC TGCCATGATT ACTATTGGGT	13560
	CAATGGTTTC ATGAGGTTTG GCACCATGAC CACCACGACC TTTAAATGTG ACGCTAAATT	13620
15	CATCTGGAGA GGCCATGATT GCGCCGCGAC GTGAATGAAT AGTTCCAGTA GGATAACCAC	13680
	TCCATAAATG TGTACCGTAA ATTCTATCTA CATTTTCCAG ACATCCAGCA TCTATCATTT	13740
	CTTGAGAACC ACCTGGCATG ATTTCTTCAC CGTACTGGAA TATTAATACA ACATTACCTT	13800
20	CTAATAAATG TTTATGTTCA TCTAAAATCT CTGCTACAGT AAGTAAAATT GCTGTATGAC	13860
	CATCATGCCC ACACGCATGC ATACATCCTG GATTTTTAGA CTTATAAGGC ACATCGTTTA	13920
	ATTCCTCGAC AGGTAACGCA TCAAAGTCAG CTCTTAATGC AATGGTAGGT CCTGTGCCCCA	13980
25	AGCCTTTAAA TGTGGCTTTG ATACCATGTC GGCCGATAGG AGTTTCAATA TCACAAGATA	14040
	ACTGGCTTAA TTGGTTAACA ATATAATCAT GTGTTTGAAA TTCTTCAAAA GATAACTCAG	14100
30	GATATTGGTG TAAATAACGT CTGAGTTGAA TTGTTTTATT TTCTTTATTA TTGCTAGTT	14160
	GGAACCAATC TAACACCCCTT ATCACTACTT TCTAAAATAA TGTTTATAGT ATAACATTTT	14220
	ATGAAATTAT CGTACTAAAT GATTGCTTTG AGATAITTTA TCTATGAATG ATAAGGCTTT	14280
35	CAAGTTATGT AGAATTACTG TATGATAAAG GTATTACCAA ACAATACTTA AGGGGGATTA	14340
	TATACTGTGG TTCAATCATT ACATGAGTTT TTAGAGGAAA ATATAAATTA TCTAAAAGAA	14400
	AATGTTTGT ATAATGAAAT AGATACAATT GAAGGTGCAA ACGGACCAGA AATCAAAATC	14460
40	AATGGGAAAT CATACATTAA CTTATCTTCA AATAATTATT TAGGACTAGC AACAAATGAA	14520
	GATTTGAAAT CaGctGCAAA AGCAGCTATT GATACACATG GTGTAGGTGC AGGCGCTGTT	14580
	CGTACAATCA ATGGTACATT AGATTTACAC GACGAATTAG AAGAAACACT AGCAAAATTT	14640
45	AAAGGAACAG AAGCTGCAAT AGCTTATCAA TCAGGATTTA ATTGTAATAT GGCTGCTATT	14700
	TCAGCTGTCA TGAATAAAAA TGATGCTATT TTATCAGATG AGCTTAATCA TGCATCAATT	14760
	ATTGATGGAT GTCGCTTATC TAAAGCTAAA ATTATTCGAG TTAACCATTG AGACATGGAT	14820
50	GATTTACGTG CGAAAGCAAA AGAAGCAGTT GAATCAGGTC AATACAATAA AGTGATGTAT	14880

55

ATTGCAGAAG AATTGGTTT ATTAACCTTAT GTTGACGACG CTCATGGTTC AGGTGTTATG 15000  
 GGTAAGGCG CTGGTACGGT TAAACATTTT GGTTTACAAG ATAAAATCGA TTTCCAAATA 15060  
 5 GGTACGCTTT CTAAAGCAAT TGGTGTCTGTT GCGGGTTATG TAGCAGGTAC AAAAGAGTTA 15120  
 ATAGATTGGT TAAAGCACA ATCACGACCA TTCTTATTCT CTACATCATT AGCACCTGGG 15180  
 GATACCAAAG CAATAACTGA AGCAGTTAAA AAGTTAATGG ATTCAACTGA ATTACATGAT 15240  
 10 AAATTATGGA ACAATGCACA ATATTTAAAA AATGGATTGT CAAAATTAGG ATATGATACA 15300  
 GGTGAGTCAG AACTCCAAT TACACCAGTA ATTATTGGTG ATGAAAAAAC AACTCAAGAA 15360  
 TTTAGTAAGC GTTTAAAAGA CGAAGGTGTC TATGTGAAAT CTATCGTTTT CCCAACAGTA 15420  
 15 CCAAGAGGTA CAGGACGTGT AAGAAATATG CCTACAGCTG CACATACAAA AGACATGTTA 15480  
 GATGAAGCAA TTGCGGCTTA TGAAAAAGTA GGAAAAGAAA TGAAGTTGAT TTAATATTTA 15540  
 20 TTTATTCCCA CGGCAAATAT TGTCGTGGGC TTTTTTTAAT GTTTAGTTTA TTAACAGT 15598

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 661 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

AAGTAAATCA ACTTACTGGG ATAAGAATAA AGGCGATTAT AGTAACAAGT TGATTTTATT 60  
 CGAAAAACAT TTTGAACCGG TTCTGGGTAT CAAGATGCAA CATAGTGGAG GTCATAGCTT 120  
 35 TGGCCACACG ATTATTACGA TTGAAAGTCA AGGAGATAAA GCAGTTCATA TGGGTGATAT 180  
 ATTCCCAACT ACTGCACATA AAAATCCTCT ATGGGTAACG GCATATGATG ATTATCCTAT 240  
 GCAATCGATT CGTGAAAAAG AACGCATGAT ACCATATTTT ATTCAGCAAC AATATTGGTT 300  
 40 CTTGTTTTAT CATGATGAAA ACTACTTTGC TGTAATAAC AGCGATAATG GTGAAAACAT 360  
 AGATGCATAT ATTTTACGTG AACATTAGT TGATAATAAC TAAAATAAAG ATGTATTACT 420  
 AAACAAATTT TCAAAAATAA AAAATTGAGC CACATCCAAT CTTACTAATT AGGGTGTGGC 480  
 TCATTTTTAA GTTTTACgAT CCAAATCAAA TATGGaTAAA ATTCgTATTA ACGCTCTACa 540  
 ATGtTAATGA CTTCAACAGT ATATGCATCT GCATAAAAAT CATAATGAAT ATTTTGACCA 600  
 50 TTTTAAATAG TTGTAATTCC ACCTTGATAA ACTAAACGGT ATTTATCAGT TTCAGGATGA 660  
 A 661

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5738 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

10 GCAGACGGTA CAGCAGTTAA AGTCGCACCA AaACTGTAGT GAATcTAATC GGTGcATTCT 60  
 TTTTAGGATT AGTTGTCGCG CTTATATATA TCTTCTTCAA AGTAATTTTC GATAAGCGAA 120  
 15 TTAAAGATGA AGAAGATGTA GAGAAAGAAT TAGGATTGCC TGTATTGGGT TCAATTCAAA 180  
 AATTTAATTA AGGATGGTTG CTAATTATGT CAAAAAGGA AAATACGACA ACAACACTAT 240  
 TTGTATATGA AAAACCAAAA TCAACAATTA GTGAAAAGTT TCGAGGTATA CGTTCAAACA 300  
 20 TCATGTTTTT AAAAGCAAAT GGTGAAGTAA AGCGCTTATT GGTTACTTCT GAAAAGCCTG 360  
 GTGCAGGTAA AAGTACAGTT GTATCGAATG TAGCGATTAC TTATGCACAA GCAGGCTATA 420  
 AGACATTAGT TATTGATGGC GATATGCGTA AgcCAACACA AACTATATT TTTAATGAGC 480  
 25 AAAATAATAA TGGACTATCA AGCTTAATCA TTGGTCGAAC GACTATGTCA GAAGCAATTA 540  
 CGTCGACAGA AATTGAAAAT TTAGATTTGC TAACAGCTGG CCCTGTACCT CCAAATCCAT 600  
 CTGAGTTAAT TGGGTCTGAA AGGTTCAAAG AATTAGTTGA TCTGTTTAAT AAACGTTACG 660  
 30 ACATTATTAT TGTCGATACA CCGCCAGTTA ATACTGTGAC TGATGCACAA CTATATGCGC 720  
 GTGCTATTAA AGATAGTCTG TTAGTAATTG ATAGTGAAAA AAATGATAAr AATGAAGTTA 780  
 AAAAAGCAAA AGCACTTATG GAAAAAGCAG GCAGTAACAT TCTAGGTGTC ATTTTGAACA 840  
 35 AGACAAAGGT CGATAAATCT TCTAGTTATT ATCACTATTA TGGAGATGAA TAAGTATGAT 900  
 TGATATTTCAT AACCATATAT TGCCTAATAT CGATGACGGT CCGACAAATG AAACAGAGAT 960  
 40 GATGGATCTT TTAAACAAG CGACAACACA AGGTGTTACA GAAATCATTG TAACATCACA 1020  
 TCACTTACAT CCTCGATATA CCACACCTAT AGAAAAAGTG AAATCATGTT TAAACCATAT 1080  
 TGAAAGCTTA GAGGAAGTAC AAGCACTAAA TCTAAAGTTT TATTATGGTC AGGAAATAAG 1140  
 45 AATTACCGAT CAAATCCTTA ATGATATTGA TCGAAAAGTT ATTAACGGTA TTAATGATTC 1200  
 ACGCTATTTA CTAATAGAAT TTCCATCAAA TGAAGTTCCA CACTATACTG ATCAATTATt 1260  
 TTTCGAATtA CAGAGTAAAG GCTTTGTACC GATTATTGCA CATCCAGAGC GGAATAAAGC 1320  
 50 AATAAGTCAA AACCTTGACA TACTATACGA TTAAATTAAC AAAGGTGCTT TAAGTCAAGT 1380  
 GACAACGGcG TCATTAGCGG GTATTCCGG TAAAAAATT AGAAAATTAG CAATTCAAAT 1440

55

GTTCTTAATG AAAGACTTAT TTAATGATAA GAAATTACGT GATTATTATG AAGATATGAA 1560  
 CGGATTTTATT AGTAATGCGA AGTTAGTTGT TGATGATAAA AAAATTCCTA AACGAATGCC 1620  
 5 ACAACAAGAT TATAACAGA AAAGATGGTT TGGGTTATAA ACAGCAAATG AGGGGTTTTA 1680  
 TGGCACATTT ATCTGTGAAA TTGCGGCTTT TAATACTAGC ATTAATCGAT TCACTGATAG 1740  
 TGACATTTTC AGTATTCGTA AGTTATTACA TTTTAGAACC GTATTTCAAA ACATATTCTG 1800  
 10 TCAAATTATT AATATTGGCA GCTATATCAC TATTCATATC GCATCATATT TCaGCATTTA 1860  
 TTTTAAATAT GTATCATCGA GCGTGGGAAT ATGCCAGTGT GAGTGAATTG ATTTTAATTG 1920  
 TTAAAGCTGT GACGACATCT ATCGTTATTA CGATGGTGGT CGTGACAATT GTTACAGGCA 1980  
 15 ATAGACCGTT TTTTAGATTG TATTAAATTA CTTGGATGAT GCACCTTGATT TTAATAGGTG 2040  
 GCTCAAGGTT ATTTTGGCGT ATTTATCGGA AATACCTTGG AGGTAAGTCA TTTAATAAGA 2100  
 20 AGCCAACTTT AGTTGTTGGT GCTGGTCAAG CAGGTTCAAT GCTGATTAGA CAAATGTTGA 2160  
 AAAGTGACGA AATGAACTT GAACCGGTAT TAGCAGTCGA TGATGACGAA CATAAACGCA 2220  
 ATATCACAAT TACTGAGGGT GTAAAAGTCC AAGGTAAAT TGCGGATATT CCAGAACTAG 2280  
 25 TGAGGAAATA TAAGATTAAA AAAATCATCA TTGCAATTCC AACTATTGGT CAAGAGCGTT 2340  
 TGAAAGAAAT TAATAATATT TGCCATATGG ATGGCGTTGA GTTATTGAAA ATGCCAAATA 2400  
 TAGAAGACGT CATGTCTGGT GAGTTAGAAG TGAACCAACT TAAAAAGTT GAAGTAGAAG 2460  
 30 ATTTACTAGG CAGAGATCCT GTTGAATTAG ATATGGATAT GATATCAAAT GAATTGACGA 2520  
 ATAAACTAT TTTAGTTACG GGTGCAGGTG GTTCAATAGG ATCAGAAAT TGTAGACAAG 2580  
 TTTGTAATTT CTATCCAGAA CGTATTATTC TACTTGGCCA TGGTGAAAAC AGTATTTATT 2640  
 35 TAATCAATCG TGAATTGCGA AATCGCTTCG GwAAAAATGT TGATATCGTT CCTATTATAG 2700  
 CGGATGTGCA AAATAGAGCG CGTATGTTTG AAATTATGGA AACGTATAAA CCATACGCAG 2760  
 TTTATCATGC AGCAGCACAC AAGCACGTGC CGTTAATGGA AGACAACCCT GAAGAAGCAG 2820  
 40 TACGTAATAA TATTTTAGGT ACGAAAAATA CTGCTGAAGC TGCTAAAAAT GCAGAGGTAA 2880  
 AGAAATTCGT TATGATTTCT ACGGATAAAG CCGTTAATCC GCCTAATGTC ATGGGAGCTT 2940  
 45 CAAAGCGAAT TGCAGAAATG ATTATTCAAA GTTTAAATGA TGAAACGCAT CGAACAAATT 3000  
 TTGTTGCAGT GAGATTTGGT AATGTACTTG GATCGAGAGG ATCTGTGATT CCACTTTTCA 3060  
 AAAGTCAAAT TGAAGAAGGT GGGCCAGTTA CTGTGACACA TCCTGAAATG ACACGTTACT 3120  
 50 TTATGACAAT TCCTGAAGCT TCTAGACTAG TTTTGCAGGC AGGGGCATTA GCAGAAGGTG 3180  
 GCGAAGTATT TGTGCTAGAT ATGGGAGAAC CAGTGAAAAT TGTAGATTTG GCACGTAATT 3240

55

	CCGGCGAAAA	AATGTTTGAA	GAGCTTATGA	ATAAAGATGA	GGTTCATCCT	GAACAAGTAT	3360
	TTGAAAAAAT	TTATCGTGGC	AAAGTACAAC	ATATGAAATG	TAATGAAGTT	GAAGCGATTA	3420
5	TTCAAGACAT	CGTCAATGAC	TTTAGTAAAG	AAAAAATTAT	TAAGTATGCC	AATGGCAAAA	3480
	AGGGAGATAA	TTATGTTTGA	TGACAAAATT	TTATTAATTA	CTGGGGGCAC	AGGATCATTG	3540
	GGTAATGCTG	TTATGAAACA	GTTTTTAGAT	TCTAATATTA	AAGAAATTCG	TATTTTTTCA	3600
10	CGCGATGAGA	AAAAACAAGA	TGACATTCGA	AAAAAATATA	ATAATTCAAA	ATTAAAGTTC	3660
	TACATTGGTG	ATGTGCGTGA	TAGTCAAAGT	GTAGAAACAG	CAATGCGAGA	TGTTGATTAC	3720
	GTATTCCATG	CAGCAGCTTT	AAAACAAGTG	CCGTCAATGT	AATTCCTTCC	AGTTGAGGCA	3780
15	GTGAAGACAA	ATATTATTGG	TACAGAAAAT	GTCTTACAAA	GTGCTATTCA	TCAAAATGTT	3840
	AAAAAAGTCA	TATGTTTATC	TACAGATAAG	GCAGCGTATC	CTATTAATGC	TAGGGGTATT	3900
20	TCAAAAGCAA	TGATGGAAAA	AGTATTCGTA	GCCAAATCAA	GAAATATTCG	TAGTGAACAA	3960
	ACGCTTATTT	GTGGTACAAG	ATACGGTAAT	GTGATGGCTT	CAAGAGGATC	AGTAATACCT	4020
	TTGTTTATCG	ACAAAATCAA	AGCTGGAGAA	CCTTTAACGA	TTACAGATCC	TGATATGACA	4080
25	AGATTTTTTA	TGAGCTTAGA	AGATGCGGTA	GAAGTATGTT	TTCATGCATT	TAAGCATGCA	4140
	GAGACAGGAG	ATATTATGGT	TCAAAAAGCA	CCAAGCTCAA	CGGTAGGGGA	TCTTGCGACC	4200
	GCATTATTAG	AATTGTTTGA	AGCTGATAAT	GCAATTGAAA	TCATTGGTAC	GCGACATGGA	4260
30	GAGAAAAAAG	CAGAAACATT	GTTGACGAGA	GAAGAATACG	CACAATGTGA	AGATATGGGT	4320
	GATTATTTTA	GAGTGCCGGC	AGACTCCAGA	GATTTAAATT	ATAGTAATTA	TGTTGAAACC	4380
	GGTAACGAAA	AGATTACGCA	ATCTTATGAA	TATAACTCCG	ATAATACACA	TATTTTAACG	4440
35	GTGGAAGAGA	TAAAAGAAAA	ACTTTTAACA	CTAGAATATG	TTAGAAACGA	ATTGAATGAT	4500
	TATAAAGCTT	CAATGAGATA	GGAGAGATTG	ACGTTGAATA	TTGTAATTAC	AGGAGCAAAA	4560
40	GGTTTTGTAG	GAAAAAACTT	GAAAGCAGAT	TTAACTTCAA	CGACAGATCA	TCATATTTTC	4620
	GAAGTACATC	GACAACTAA	AGAGGAAGAA	TTAGAGTCAG	CATTGTTGAA	AGCAGACTTT	4680
	GTCGTGCATT	TAGCGGGTGT	TAATCGACCT	GAACATGACA	AAGAATTCAG	CTTAGGAAAC	4740
45	GTGAGTTATT	TAGATCATGT	ACTTGATATA	TTAACTAGAA	ATACGAAAAA	GCCAGCGATA	4800
	TTATTATCGT	CTTCAATACA	AGCAACACAA	GATAATCCTT	ATGGTGAGAG	TAAGTTGCAA	4860
	GGGGAACAGC	TATTAAGAGA	GTATGCCGAA	GAGTATGGCA	ATACGGTTTA	TATTTATCGC	4920
50	TGGCCAAATT	TATTCGGCAA	GTGGTGTAAG	CCGAATTATA	ACTCAGTGAT	AGCAACATTT	4980
	TGTTACAAAA	TTGCACGTAA	CGAAGAGATT	CAAGTTAATG	ATCGGAATGT	TGAACTAACG	5040
55							

ATTGAAAATG GTGTACCTAC AGTACCAAAC GTATTTAAAG TGACATTGGG AGAAATTGTA 5160  
 GATTTATTAT ACAAGTTCAA ACAGTCACGT CTCGATCGAA CATTGCCGAA ATTAGATAAC 5220  
 5 TTGTTTGAAA AAGATTTGTA TAGTACGTAT TTAAGCTATC TACCTAGTAC aGACTTTAGT 5280  
 TAYCCCTTAC TTATGAATGT GGATGATAGG GGTTCCTTTA CAGAATTTAT AAAAACACCG 5340  
 GATCGTGGTC AAGTTTCTGT AAATATTTCT AAACCAGGTA TTACTAAAGG TAATCACTGG 5400  
 10 CATCATACTA AAAACGAAAA ATTTCTAGTC GTATCAGGTA AAGGGGTAAT TCGTTTTAGA 5460  
 CATGTTAATG ATGATGAAAT CATTGAATAT TATGTTTCTG GCGACAAATT AGAAGTTGTA 5520  
 GACATACCAG TAGGATACAC ACATAATATT GAAAATTAG GCGACACAGA TATGGTAACT 5580  
 15 ATTATGTGGG TGAATGAAAT GTTTGATCCA AATCAGCCAG ATACGTATTT CTTGGAGGTA 5640  
 TAGCGCATGG aAAAACGTAA TTAATGACA ATAGTTGGTA CAAGGCCTGA AATCATTCGT 5700  
 20 TTATCATCAA CGATTAAAGC ATGTGATCAA TATtTTAA 5738

## (2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9062 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

ATCATCAACA AGAATGATAT TTTCCCATC TACTATATCT TTTACCGCAG ATAACTTCAC 60  
 TCTCACACCT TGCTCACGTA ATTCTTGAGT TGGTTGAATA AATGTTCTTG CAACATATTG 120  
 35 ATTTTTAACT AGTCCCATTT CATATGGCAA ACCTATTTCT TCAGCATAAC CACTCGCAGC 180  
 TGATAGCGAT gAATTGGGTA CACCGATGAC CATATCAGCA TTTACAGGGC TTTCTTGGGC 240  
 TAATTTTTTA CCAGAAGCTT TACGTACTGC ATGGACATTT TTACCAGCTA TTGTTGAGTC 300  
 40 TGGTCTAGCA AAATAAATAT ATTCCATCGC AGAAATTGCA GTTGTCGTAT GATGTGTATA 360  
 AGATTTAACT GTAATACCTT TATCGTTAAT CACGACATAT TCACCTGCAT GAATATCTTG 420  
 AACAAATTCT GCACCTAACA CATCTATTGC ACATGTTTCA CTTGCAAGGA TGTATGTCCC 480  
 ATCTTTCATT TTACCTACAA CAAGTGGTCT GATAGCATTT GGATCTACTG CGCCATATAA 540  
 CGCATCTTTA GTTAAATCG CAAATGTAAC ACCGCCTTTA ACTTTTCGCA AACTTTCTTT 600  
 50 CAACGCTTCC TCAAAAGTAG GAGCTTTACT TCGACGTATC AAATGCATAA TGAATTCAGT 660  
 ATCAGAAGAC GAATGGAAGA TAGCACCTTG TTTTCTAAA TTCTGACGCA ATGATTTAGC 720

55

	CGGTTGAATA TTTTCAATAC CTTTATTACC TGAAGTAGCA TAACGGACGT GACCAATTGC	840
	ATGTTGATAT CCTTTTAATC GTTCCATTTG ATCATCTTTA ATCGCTTCAG TTAGTAAGCC	900
5	TAATCCTCGC TCGCCTTTTA ATTCATTTTG ATCAGAAACA ACTATACCTG cACCTTCTTG	960
	ACCACGATGT TGCAAACTAT GAAGTCCCAT ATAtGTTAGT TGCGCTGCTT CaGGATGATT	1020
	CCAAATACCA AACACGCCAC ATTCTTCGTT TAATCCTGAG TAGTTAAACA TTGaGCAATT	1080
10	GCCCCtTCCC ATATTTGTTT AATATCTGAA ACATTTTCAC TAATCTCTGT aTATGGTGTT	1140
	GTTACCTTGx aATTATCACT ATCTGTTAAA AGTCCAATT CTATTGCATT ATCAATATTT	1200
15	AAAGTTTTAC CTGATTTAAC AGAAACAACA TATCGGCCTT GCGTCTCACT AAACAATTGT	1260
	GCATTTGTTA TATCTATTGA AGATTTTAAT CCTAAACCGT AATGCGCACT TAGTTTAGCT	1320
	AAGTAATCA GTAAGCCACC TTTACCAACT GTTGAACAT GTGATAATAG TCCTTCACGA	1380
20	ATAGCGGTCT TGATTGATTC ACCTTTTCA ACTTCTGAAC TCAAATCTAA TGA CTCAAAT	1440
	TCATGATTAA CTTTGCCATA AATTAACTTT TCAAGTTGAC TACCACCAA GTCGTCCTTA	1500
	GTATCACCGA TTAAATATAA TTTATCTCCA ACTTGAGGT CAAATCATT TAAATAATTT	1560
25	ACATTTTCAA TCAAACCTAC CATTCCAACA ACTGGTGTG GGAAAATAGA AGTACCTTTC	1620
	GTTTCGTTAT ATAAAGATAC ATTACCAGAA ACTACTGGTG TCTTAAGAAT GTCGCATGCT	1680
	TCTGCCATAC CTTTCGTTGA ATCTATCAAC TGTGATAGA TTTCTTTCTT TTCAGGAGAA	1740
30	CCATAATTTA AACAATCTGT CATTGCTAAT GGTGTTGCAC CCACGGCAAT TAAATTCGA	1800
	TAAGCTTCAG CTA CTACCAT CTTTCCACCT TCATATGGAT TGTTATATAC ATAACGCGCT	1860
	TCACCATCAA TTGTTGAAGC AATTGCCTTA TTTGTGCCTT CCACACGTAC TACCGATGCT	1920
35	TGAAGTCCTG GCTTAATTAT CGTATTGGCA CCAACTTGTT GGTCGTATTG ATCATATAAA	1980
	TAGTGTTTAG ATGCTATAGT CGGATGCTTA AGTAATTTAA AGAAAGTATC TTTAACATCG	2040
40	ATGTGTGTAT AATCATTTTT AGAAGTATTA TAATCTTTTT CTTCTCCTTC TAAAATATAT	2100
	ACAGGTGCTT CATCAGCTAG TGGTTCAACT GGAATGTCAG CATAAACTTC GTCATCATAT	2160
	GTTAAAACAA AACGATTTGT ATCTGTAAC TCACCTATAA CAGCACTATC CAATTCGTGC	2220
45	TTATCAAATA AATCTAAGAA TTTTGTTC GTACCTTTTT CAACAACTAG TAACATACGT	2280
	TCTTGAGTTT CTGAAAGCAT CATTTCATAA GGAGAAATAC CTGGCTCACG TGTTGGCACT	2340
	TGTTCTAATC TCAAATGTAA CCCACTACCA CCTTTGCCG CCATTTCAGA CGATGAAGAT	2400
50	GTTAAACCAG CAGCACCCAT ATCTGAATA CCAACTAATT CATCAAATGT AATTGCTTCA	2460
	AGTGTGCTT CCATTAATTT TTTACCTACA AATGGATCAC CGATTTGTAC AGAAGGTCGT	2520
55		



	CGACCAGTTT TCAAACCAAC ATAAATGACC GAATTACCTA CACCTTTTGC TGTGCCTTTT	2640
	TGAATCATGT CGTGATTGaT AACACCAACA CACATTGCAT TAACAAGTGG ATTGCCATCA	2700
5	TAACGTTTCAT CAAATTCGAT TTCACCAGCA GTTGTGGaA TACCAATGCA GTTACCATAA	2760
	CCTCCGATAC CCTTTACAAC ACCTTTAAGT AATCTTTGGT TTTGTTTATT ATCTAATTCT	2820
	CCAAATCTAA GACTGTTTAA CAAATTAATA GGTCTAGCCC CAATAGAGAC AATGTCACGA	2880
10	ATGATTCCAC CAACGCCTGT AGCAGCCCCT TGATATGGTT CAATTGCTGA TGGATGATTG	2940
	TGAGACTCTA CTTTAAATAC TACGGCTTGA TTATCACCTA TATCGACTAC CCCTGCACCT	3000
	TCACCAGGCC CCATAAGCAC ATGGTcACCT GACGTAGGAA ATTGCTTTAA AAACGGTTTA	3060
15	GAATGTTTAT AAGAGCAATG TTTACTCCAC ATAACAGAAA AGATACCTGT TTCTGTAAAG	3120
	TTAGGTTGTC TGCCTAAAAT ATCGCAAAC TTTTCATATT CTTGATCaCT TAATCCCAT	3180
20	TCTTGATATA CTTTTCAAG TTTAATTTCT TCAACGCTTG GTTCGATAAA TTTAGACATG	3240
	TTGTTCCCTC CAACTTTTTA CCATCGCTTC AAATAATTTT ACACCACTAT CAGTACCTAA	3300
	CAACGTTTCT AAAGCTCTTT CagGATGtGG CATCATGCCA CATACTTGC CTTTTTCGTT	3360
25	AACAATTCCT GCAATATCAT CATATGAACC GTTCGGATTA TTCACATATT TCAGAATAAT	3420
	TTGATTGTTA GCTTTTAATT GTTGATATAT TTCATCAGTA CAATAATAAT GACCTTCACC	3480
	GTGAGCTACA GGATATATAA CTTTTTCACC TTGTTCATAA AGATTTGTAA ATGCCGTTTG	3540
30	ATTATTCACT ATTTCTAACT CTTCAATTCT ACTAATAAAT AAATGTGAAT CGTTATGCAA	3600
	TAATGCACCA GGTAATAAGC CTATTTCACT TAAAATTTGA AACCATTAC AAACACCTAA	3660
	TACTGGCTTA CCTTCAGCTG CAAGACGTTT AACTTCCGAA ATAATCGGsG CTACACTAGC	3720
35	CATTGCCCCA GATCTTAAGT AATCCCCGAA TGAAAATCCA CCAGGAATAA GTACGCCATC	3780
	AAATēCACTT AGTGATGTTT CTCTATAATC TACATATTCC GCTTCAACAC CACTTTTAAT	3840
40	AGCAGCATT A ACATGTCTC TATCACAATT CGAACCTGGA AAAACAAGAA CCGCAAATTT	3900
	CATTTTATGC ATTCTCCTTT TCATCATCTA ACACTTTATA GCTATATTCT TCAATCACTG	3960
	TATTTGCAAA CAATTTTCA CTTAGAGTTG TAATAATGTT GTGTACCTTT TCATCACTAA	4020
45	CCTCATCCAC TGTCATATAT AATACTTTTC CTACACGAAT ATCATTCACT TGTGCATAAC	4080
	CTAAGTCATG TACAGCTCGA GTAAGCGTTT GTCCTTGCGT ATCTAATACT TGTGGTTGTA	4140
	ATGTGATATG TAGTTCAATT GTTTTCATTA TTTTAAATCC TCCAATTTGT TTAAAAATAT	4200
50	TTGATATGTT TCAATCAGTG ATCCAGTGTT ATTTCTATAT ACATCTTTAT CAAAGTTTGC	4260
	ATGGTAGCT TTATCCCAA TTCGACATGT ATCTGGAGAT ATTTTCATCCG CTAACAAAAT	4320
55		

ATCCATTAAT TGTTCACA CATTATTAAT CTTTAATGCT TTGGATTTTA GTATTTCAAT 4440  
 ATCTTCATCT GATGCTATAT TGAGCAATTT AACATGGTCA TCCGTTATCA ACGGATCATT 4500  
 5 TAACGCATCA TTTTATAGA AAAATTCCTAC AAGTGGTTCT CTAAAACTT CACCATTTTC 4560  
 AAAACCTAAA CGCTTTGTAA TAGATCCACT AGCAATATTA CGAACAACTA CTTCTAATGG 4620  
 AATTATTTTC ACAGGCTTAA CTAATGTTC TGTTCAGAT AATTGTTTAA TAAAGTGAAT 4680  
 10 TTCTATTCCA TTTCTTGTA AATATTTAAA TATAATAGAA GTAATTTGAT TATTTAATCG 4740  
 CCCCTTACCT GCCATTGTGT CTTTCTTAGC CCCGTTTCCA GCAGTAACTT CATCTTTATA 4800  
 TTCAACTCTT AATTCATTTT CTTGATTGTG TGAGAAAATG CGCTTCGCTT TTCCTTCATA 4860  
 15 TAATAATGTC ATGCTTTAAT TACTCCCCTC AAATTTAGCG TACATATCTT GTTCAGTTTG 4920  
 GTTACATCA TTCGTTAGTA CAGTCATATG CCCCATTTTT CTGCTATCTT TACGCTCAGA 4980  
 20 CTTACCATAA ATATGTAAGT GCCACTCTGG ATGTTCAATTA AATTCATTTT CCAATAAATC 5040  
 TAAATCTTTA CCTAGTAAGT TCATCATGAC TGCTGGCTTT AATAATTCAA TTGAATTTGG 5100  
 TAATGATTGT CCGGTAAGT CTAAAATATG AGTATCAAAT TGTGAATAAT CACATGCTTC 5160  
 25 AATTGAATAA TGTCCGGAAT TGTGAGGCTT TGGTGCTATC TCGTTCACAT ACAATTGGTT 5220  
 GTTACTATCT ATAAAAAATT CAACTGTAAA TGTTCGAATG AAATGAATCG ATTGGATAAT 5280  
 TTTATTAAT TGTCTTTTCG CCTCAGCTGT TTTATCTATT CTCGCTGGAA CAATTGTTTT 5340  
 30 GAAAAGTATT TGATTTCTAT GCTCATTTTC TTGTAATGGG AAAAAAGTGA TTTGATTGTT 5400  
 GTTCTCTCTT GTAACAGTAA GAGATACTTC TTTCTTGATA TTCAAATATT TTTCAGCTAC 5460  
 GCATTCATA GTTTCAATTA ATTTAAAACC TTCTTGTAAG TCTTTTTCGT TGTTAATTAA 5520  
 35 AACTTGACCT TTGCCATCGT AGCCACCAAA TCTAGTTTTT ACAATAAAAG GATATCCTAA 5580  
 TGTTCGAAT GCTTTGTCAA TATCTGTAGA TTCTTTTACT GAAATGAACG GGACAACTTT 5640  
 GGTACCAGCA CTTTTTAATG TTTCTTTTTC AGTTAAGCGA TCTTGTAATA ACTGTATAGC 5700  
 40 TTGGTAACCT TGCAGGAATAT TGTACTTTTC ACATAATAGT TTTAATTGTT GGGCTGAAAT 5760  
 GTTTTCAAAT TCATAAGTAA TCACATCACA TTTTGTCTCT AATTGATTGA GTGCCTTTTC 5820  
 45 ATCGTCATAC TTGGCTTGTA TAAATTCGTG TGCAACGTAT CTACATGGAC AATCTTCAGA 5880  
 AGGATCCAAT ACAACCACTT TATAACCCAT TTTTGTAGCT GATTGTGCCA TCATCTTTCC 5940  
 AAGCTGACCA CCACCAATAA TGCCAATAGT CGCACCAAAC TTAAATTTAT TGAAGTTCAT 6000  
 50 TTTGCATGTC CTCCACTTTT TGAATTAACG AAGATTCTTA CTGATTAGT TTTTCAACTA 6060  
 AAGAAGGATT TTGAATACTT AACATTCTTG CTGCAAGTAT ACCTGCGTTT TTAGCACCTG 6120

55

EP 0 786 519 A2

	AAGAATCTAT ACCCTTTAAA CTTTTTGTTC CAATCGGCAC TCCAATAACT GGTAGCGTCG	6240
	TTAATGATGC AACCATACCT GGTAAATGTG CCGCACC GCC AGCGCCTGCA ATGATAATGT	6300
5	TTATACCTCT TTCTCTCGCT TCAGAAGCAA ATTGAACCAT CATTTTGGC GTACGATGTG	6360
	CGGATACTAC TTGTTTTTCG TACGGAATTT CAAAATAATC CAACATGTTA CAACTCTCTT	6420
	GCATAATTTT CCAATCGGAA GAACTGCCCA TAATGACTGC TACTTTCAC TGTACACCC	6480
10	TTTCAAAAGT TTGAATTGTG AATTACTTTA GTTGTATATT ATAGATATAG CATAACAAGC	6540
	AATTTCTGCT TTTTCAATCA AAAATCGAAC TTTATTTTGA TTTTTTATTT GAATTTACGT	6600
	CTTTTGCTAT GTAAATTAGT TTTATAAACT AACAAAGTTA GGATATTGAC AATAGGAGGA	6660
15	GAAGTTTTTA TGGTTGCTAA AATTTTAGAT GGTAAACAAA TTGCCAAAGA CTACAGACAG	6720
	GGGTTACAAG ATCAAGTTGA AGCGCTAAAA GAAAAGGGTT TTACACCTAA ATTATCCGTT	6780
20	ATATTAGTTG GTAATGATGG CGCTAGTCAA AGTTATGTTA GATCAAAAAA GAAAGCAGCT	6840
	GAAAAAATTG GTATGATTCT AGAAATCGTA CATTTGGAAG AAACAGCTAC TGAAGAAGAA	6900
	GTATTAAACG AACTAAATAG ACTAAATAAT GATGATTCTG TAAGTGGTAT TTTGGTACAA	6960
25	GTACCATTAC CAAAACAAGT TAGCGAACAG AAAATATTAG AAGCAATCAA TCCTGAAAAA	7020
	GATGTGGACG GTTTTCATCC AATAAATATA GGGAAATTAT ATATCGATGA ACAAACTTTT	7080
	GTACCTTGCA CACCGCTCGG CATCATGGAA ATATTAAAAC ATGCTGATAT TGATTTAGAA	7140
30	GGTAAAAATG CAGTTGTAAT TGGACGAAGT CATATTGTCG GACAACCACT TTCTAAGTTA	7200
	CTACTTCAA AAAATGCATC AGTAACAATC TTACATTCTC GTTCAAAGA TATGGCATCA	7260
	TATTTAAAAG ATGCTGATGT CATTGTCAGT GCAGTTGGTA AGCCTGGTTT AGTAACAAAA	7320
35	GATGTGGTCA AAGAAGGAGC AGTAATTATC GATGTTGGCA ATACGCCAGA TGAAAATGGC	7380
	AAATTAAAAG GTGACGTTGA TTATGATGCG GTTAAAGAAA TTGCTGGAGC TATTACACCA	7440
	GTTCTGGTG GCGTTGGTCC ATTAACAATT ACTATGGTAT TAAATAATAC TTTGCTTGCA	7500
40	GAAAAAATGC GTCGAGGTAT TGATTCGTAA AGAGCCTGAG ACATAAATCA ATGTTCTATG	7560
	CTCTACAAAG TTATAATGGC AGTAGTTGAC TGAACGAAAA TTCGCTTGTA ACAAGCTTTT	7620
45	TTCAATTCTA GTCAACCTTG CCGGGGTGGG ACGACGAAAT AAATTTTACG AAAATATCAT	7680
	TTCTGTCCCA CTCCCTAATA ACTGAGTTTT AATGAAGTCT TTTAACCCAC ATTAAATATT	7740
	ATTTTGCAAT TGCAATGAAT AACAAGAAAA ATCTGGGACA TTAATCGATC AAATGCTCCC	7800
50	TTCAAAGTAG ACATTGAATA AATGAAGGCT TTGAAGGGAG CATTTCACTT TGTACTTGGC	7860
	TCAACAATTT TATATAGACA GTAGTTAATT GAATGAAAAT AAGCTTGTA CAAGTTTTCA	7920

55

GTTGGGGATG GGCCCCAACA CAGAAGCTGT GACTATGATA AAGTACTACT ACATAGTTAA 8040  
 TCATTAGTGG TTCTTTATCA TTTTCGCCTC CCTTTTCTTA TTGTTTGTGAT ACACAAAAAT 8100  
 5 TTAAGTTCAA ACTGTGCAAT AAAGTTATAT TTGATTTCOA ATTATCCCTA AATTATTAAT 8160  
 TKTACAATTG TGGCAGATTT TCAAAATAAT AATTATTTCC TCATTATTTA TAAATTTATA 8220  
 TTTAAATTTT ATTCTTTATA GGGTAAGATT AGGACTATAG TATGATGTGT AATAATATA 8280  
 10 AATTAAGGTA TAGTAAAGCT AACTCAGAAA TGACTTATCA TTCGGAGGTT ACATTATGAA 8340  
 TAAACTATTA CAGTCATTAT CAGCCCTCGG TGTTTCTGCT AACTAGTAA CACCAAATTT 8400  
 15 AAATGCAGAT GCAACGACGA ATACTACACC ACAAATTAAG GCGCTAATG ATATCGTTAT 8460  
 TAAGAAAGGT CAAGATTATA ACCTTCTAAA CGGCATAAGT GCATTTGATA AAGAAGATGG 8520  
 AGATTTAACC GATAAAATTA AAGTCGATGG CCAAATGAT ACATCTAAAT CTGGTAAATA 8580  
 20 TCAAATTAAG TATCATGTCA CTGATTGAGA TGGTGCAATT AAAATTTCCA CTAGGTATAT 8640  
 TGAGGTTAAA TAGCCCTCAT CACTATACTG CAAATAAAAT GGTAGCAAAC GAACATGTTT 8700  
 TGCTACCATT TTATTGTGTA TTCTAACTTC ATCTGCAACT TTAACCCAAA TATTGTATTT 8760  
 25 TTTCTGTATA CCAAAGGACT ACCTATCAAA TTATTAAAC TTAAGTCTC TTTTAAAAA 8820  
 AATGTTTTGA TTTTGAACAA ACAAATTTCC ACTTTTCATT GTTTAAGCAT AAATTACTTT 8880  
 TGGCAAATTC CTTATTAAAA TGTTTGCGCT TCCTTTCAAT CAACTAGCCA TCATTTTCAA 8940  
 30 TTTATTAGAC AATTTCAAAC TTTTATTATT TTCATTCAAT TAACCTTTAA TTGAAAGCTA 9000  
 TTCTCAACTT TCCTTTTAAA TATGAAGCAA TTTTTCAAA AACGCTATTA GTCACAAAAT 9060  
 GT 9062  
 35

## (2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2738 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

AAATATTTTT TCAAACTAT GTGAAATGG aCCATGTCTA aATCATGTAA TAATGCAGyA 60  
 CATAATGCCA ACGGTCTmTC TTTATTGTCC CATGCATCAT GACCAATAAA TGACTCATCA 120  
 50 ATTAATCGTC TAACTATTTT ATACACACCT AAAGAATGTC CAAAGCGACT ATGTTCTGCT 180  
 GTGTGAAAAG ATAGGTACAG TGTTCTTAGT TGTCTAATTC GACGTAACTT TTGGAATTCC 240

55

EP 0 786 519 A2

	TCTTTAAAAA	CTTTTCTTC	TACTAATTTT	AAATCTACAT	ATGCGTTAGT	CATTATTCCC	360
	CTCCTTTTCG	TTTAATATAA	TATTTAATTT	ACTTAAAATG	CTTTGTACAT	AAGTGCTAAG	420
5	TCTAACTTTT	CGCCATACAT	TTCTGGCTCA	TAAGAGCGTA	AGATTGTAAA	ACCTTGCTCT	480
	TTATAGTAAG	CTACTGCTTC	TTCAATTTTA	TTATCTACTT	CTAAGTAAAC	ACCTTCAAAT	540
	TTATCTTCAA	AACGTGATAA	TCCTTCATTT	AACAATGCTG	TACCATAACC	TGTATGTTGC	600
10	GATTCTGGTT	TAACATAATG	AGCTGATAAA	TATAATTCTT	CACCGTAAAT	AAAGTTAGCA	660
	AAGCCAACGA	TGTCATTACC	TTCTTCAACG	ACTAAGAATA	ATTGTTCTTG	AAGTCTTTTC	720
	TTTAAATGAT	GTTCAATTATA	TGAAGCTtCT	AACAAGTGAT	TAAGTGTGT	CGCAGCGTAT	780
15	ATATTTAAGT	ATGTATTAAA	CCAAGCTTTA	GTTGCGACAT	CTCTAATTTG	AACAACATCT	840
	TTTTCAGTTG	CTTGTCTTAC	CTTGAACATG	ACTTTCTCCC	CTTATTAACA	AGTTTTAATA	900
	ACGGCATTAT	ACCACAACCT	GCTCAATACT	TAATAAACAA	TGATTGTCTA	TTCAATTTAT	960
20	ATATtTATAT	TTCCGTAA	AATTAAAAAT	AAAAAATAAC	GAAGCAAAAA	AtCACTTCGT	1020
	TTAGTATGAG	GTATGTCTTA	TTGCAATATA	CTATTCCACT	CAGTTGCACG	TGCTAAGGCA	1080
25	TAGTTGTCTT	TCATGATGTC	ACCAGGCTTT	TCAGCAGTTC	CAATAATATA	ACCATTTAAA	1140
	GTGGCACCTA	rAAAGTCTAA	ACTATATTTT	ATTTGCGTAA	TTGCTGGTTC	GCTTTTATTT	1200
	TTGGACAATC	TCCACCAACT	AAAATAACTC	TAAATCCTT	TCGGCCATT	TGTGCCTTAA	1260
30	AATTAGGATA	TCGTTTATCT	TGTAATGTTT	CTGACCAATG	TTGATAAAT	GCTTTCAATG	1320
	GTGCTGAAAT	GCTATACCAA	TACACTGGTG	ATGCAAAAAT	AATTGTATCA	CTAGCCAATA	1380
	TTTTATCTAG	AATCGGCAAA	TAGTCATCGT	CATATGAAGT	AATAGTCTCT	GCTGTATGTC	1440
35	TCACGTCACG	TATCGGTTTA	AACTGATGTT	GTGTCACGTC	AATCCATTGA	TACTCTAAAT	1500
	CTTGCAAAGC	GAATTTTGTT	AATTGTGCAG	TATTACCGTT	TGGTCTACTC	CCACCAAACA	1560
	AAACAGTAAT	CATTTTAGCC	TAACCTCACT	TTTGATTAAAT	AAATATCTGT	GTTTTTCGTT	1620
40	ACCTAATTAT	ACTATCATAA	GCTTTGCCTA	CCGAATAGTA	AAACGCTTAC	AACTTTTATA	1680
	TAAATTTGAC	GAAATTTTCGT	CATGCCTTAT	ATAACGTCGT	TTGTGATACG	GGGCTAATTC	1740
45	ATGATGAAAT	TAGATACATA	TATCACCATT	AAATACAATT	CATTTAGTCT	TCAATCGGAA	1800
	ACAGTTCATC	GATATATTGA	ATCTCATCAT	CTGATAAAAC	GATATCTGCA	GCTTTAATAT	1860
	TTTCAACGAC	TTGTTCTGCA	CGTTTTGCAC	CAGGAATAAT	CACATCGATA	GCTGGTCTCG	1920
50	TTAAATAAAA	TGCTAATACA	ATGTTTCGCAA	TTGAAGTTTG	ATGTGCTGCA	GCTATGCTTT	1980
	CCAAAGCTTT	TACGCGACGC	ACATTTTCTT	CAAATACACC	TGGTTTAAAA	TCACGACGTG	2040
55							

GCTAATGGGA AATATGGAAT AAATGTGATT TGGTGATCAA CACAATATTG TAATACTGCC 2160  
 TCATTTTCGC GATGCAATAA ATTATATTCT AACTGTACAA CATCAACGTA ACCATCTTTA 2220  
 5 TTTGCTTCTT TAAGTTGATC TAATGTGAAA TTTGATACAC CAATTGCTTT AATCTTCCCT 2280  
 TGTTCCTTAA GCTCTGTAA TGCTGCAACT GCTTGATCTT TCGGAGTGTT GTTATCCGGA 2340  
 AAATGAATAT AATATAAATC GATATAATCA GTTTGTAGAC GTTCAAACCT ATTCTCAACT 2400  
 10 TGTTGTTTTA AATATCCGG TTGATTGTTT TGATGTACTT CTTGATTTTC ATCAAATTCA 2460  
 TGAGACCCCT TCGTAGCAAT TTTAATTTGC TCTCGCGGAT ATTCTTTAAC AACTTCTCCA 2520  
 ACCAATTCTT CTGATCGTTC TGGCCCATAA ATATATGCCG TATCTAATAA ATTAATACCA 2580  
 15 TGATTAATGG CTTGACGAAC AACATCTTTT CCTTGTTCTT CATCTAAGTT CGGATATAAA 2640  
 TTATGCCCAa CCTAtGCGTT CGTCCCAAGT GCGATTGGAA ACACTTCAAC ATCAGATTTA 2700  
 20 CCTAAGTTTA CAAATTGCTn CATTAGACCC AGCnCTT 2738

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9425 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

GATTAGATGA TATTTAACGA AAATTAAGrT GmAATACTtG AATGTArGAa GTCTGATGTC 60  
 GAAAATAGCT ATTAAAAATAG AGTAGACGTA ATGtAAATGA AAGCACCTAA AATAGAAAAA 120  
 35 TTTCAAAAAT AGCGTAATTA TTATAATAAA TAGACTGCCA ATAAAATGCA ATTTTTCACCT 180  
 TATAAcATTC TTCAAAAAAT AATAGCAAAA TTATGTAAAA AATATCTTGT CATGGCAAGA 240  
 TTGGCTGTGC TATAATCTAT CTTGTGCTTA AGAACGGCTC CTTGGTCAAG CGGTTAAGAC 300  
 40 ACCGCCCTTT CACGGCGGTA ACACGGGTTC GAGTCCCGTA GGAGTCACCA TTTTTTAGGT 360  
 CTCGTAGTGT AGCGGTAAAC ACGCCTGCCT GTCACGCAGG AGATCGCGGG TTCGATTCCC 420  
 45 GTCGAGACCG TACAAATGCC TATCCAAGAG GATAGGCATT TTTTTCGTT TAATATTATA 480  
 TTAATAAAAG ATATATGGAC GAATGATAAT CATATTGATT TATCTGTTCTG TCCATTTTCT 540  
 TTAAATGTA TGAACCTCAA GTAACCTAGT GGTTGGATAT GAAAGATAAA CGTAGACAAT 600  
 50 AAAATCTTTA TTAGACGTAC AAACATATGC TACTGTCAAC ATATTCTTTC GTTGTGATAT 660  
 GCCACCAGTC CTCCATAACA TCAATTGTTA AAGTAACGAA TAACGAATAA TGATATTTAT 720

55

	GACCTCATCA TTGTGTTAAA TATCATTGTC ACAATCCGCC GTGAGAACT AATAAAAAAT	840
	AGTAATATAT AAGTTTATAT TGGAAAATAG AATTAATAGC TTATAAATGG TAAATTATAT	900
5	AATAGGTTAC TATACGTTAT AAGACGGAAA ATGCGCACAA TAACAAAAAT AGTAAGCGAC	960
	ATCCTGTGAT TTTTACACA AACATAAACG ATAAAGAACA AAAAATGATA AAATAATATT	1020
	AATGATTTAA GAAAAGAGGT TTATGCAAAT GGCTAGAAAA GTTGTGTAG TTGATGATGA	1080
10	AAAACCGATT GCTGATATTT TAGAATTTAA CTAAAAAAA GAAGGATACG ATGTGTACTG	1140
	TGCATACGAT GGTAAATGATG CAGTCGACTT AATTTATGAA GAAGAACCAG ACATCGTATT	1200
	ACTAGATATC ATGTTACCTG GTCGTGATGG TATGGAAGTA TGTCGTGAAG TGCGCAAAAA	1260
15	ATACGAAATG CCAATAATAA TGCTTACTGC TAAAGATTCA GAAATTGATA AAGTGCTTGG	1320
	TTTAGAACTA GGTGCAGATG ACTATGTAAC GAAACCGTTT AGTACGCGTG AATTAATCGC	1380
20	ACGTGTGAAA GCGAACTTAC GTCGTCATTA CTCACAACCA GCACAAGACA CTGGAAATGT	1440
	AACGAATGAA ATCACAATTA AAGATATTGT GATTTATCCA GACGCATATT CTATTAAAAA	1500
	ACGTGGCGAA GATATTGAAT TAACACATCG TGAATTTGAA TTGTTCCATT ATTTATCAAA	1560
25	ACATATGGGA CAAGTAATGA CACGTGAACA TTTATTACAA ACAGTATGGG GCTATGATTA	1620
	CTTTGGCGAT GTACGTACGG TCGATGTAAC GATTCGTCGT TTACGTGAAA AGATTGAAGA	1680
	TGATCCGTCA CATCCTGAAT ATATTGTGAC GCGTAGAGGC GTTGGATATT TCCTCCAACA	1740
30	ACATGAGTAG AGGTCGAAAC GAATGAAGTG GCTAAAACAA CTACAATCCC TTCATACTAA	1800
	ATTTGTAATT GTTTATGTAT TACTGATTAT CATTTGGTATG CAAATTATCG GGTTATATTT	1860
	TACAAATAAC CTTGAAAAAG AGCTGCTTGA TAATTTTAAG AAGAATATTA CGCAGTACGC	1920
35	GAAACAATTA GAAATTAGTA TTGAAAAAGT ATATGACGAA AAGGGCTCCG TAAATGCACA	1980
	AAAAGATATT CAAAATTTAT TAAGTGAGTA TGCCAACCGT CAAGAAATTG GAGAAATTCG	2040
40	TTTTATAGAT AAAGACCAAA TTATTATTGC GACGACGAAG CAGTCTAACC GTAGTCTAAT	2100
	CAATCAAAAA GCGAATGATA GTTCTGTCCA AAAAGCACTA TCACTAGGAC AATCAAACGA	2160
	TCATTTAATT TTAAAAGATT ATGGCGGTGG TAAGGACCGT GTCTGGGTAT ATAATATCCC	2220
45	AGTTAAAGTC GATAAAAAGG TAATTGGTAA TATTTATATC GAATCAAAAA TTAATGACGT	2280
	TTATAACCAA TTAAATAATA TAAATCAAAT ATTCATTGTT GGTACAGCTA TTTCATTATT	2340
	AATgCACAGT CATCCTAGGA TTCTTTATAG CGCGAACGAT TACCAAACCA ATCACCGATA	2400
50	TGCGTAACCA GACGGTCGAA ATGTCCaGAG GTAACATAC GCAACGTGTG AAGATTTATG	2460
	GTAATGATGA AATTGGCGAA TTAGCTTTAG CATTTAATAA CTTGTCTAAA CGTGTACAAG	2520

55

	GTGATGGTAT TATTGCAACA GACCGCCGTG GACGTATTCG TATCGTCAAT GATATGGCAC	2640
	TCAAGATGCT TGGTATGGCG AAAGAAGACA TCATCGGATA TTACATGTTA AGTGTATTAA	2700
5	GTCTTGAAGA TGAATTTAAA CTGGAAGAAA TTCAAGAGAA TAATGATAGT TTCTTATTAG	2760
	ATTAAATGA AGAAGAAGGT CTAATCGCAC GTGTAACTT TAGTACGATT GTGCAGGAAA	2820
	CAGGATTTGT AACTGGTTAT ATCGCTGTGT TACATGACGT AACTGAACAA CAACAAGTTG	2880
10	AACGTGAGCG TCGTGAATTT GTTGCCAATG TATCACATGA GTTACGTACA CCTTTAACTT	2940
	CTATGAATAG TTACATTGAA GCACTTGAAG AAGGTGCATG GAAAGATGAG GAACCTGCGC	3000
15	CACAATTTTT ATCTGTTACC CGTGAAGAAA CAGAACGAAT GATTGACTG GTCAATGACT	3060
	TGCTACAGTT ATCTAAAATG GATAATGAGT CTGATCAAAT CAACAAAGAA ATTATCGACT	3120
	TTAACATGTT CATTAAATAA ATTATTAATC GACATGAAAT GTCTGCGAAA GATACAACAT	3180
20	TTATTCGAGA TATTCCGAAA AAGACGATTT TCACAGAATT TGATCCTGAT AAAATGACGC	3240
	AAGTATTTGA TAATGTCATT ACAAATGCGA TGAAATATTC TAGAGGCGAT AAACGTGTCG	3300
	AGTCCACGT GAAACAAAAT CCACTTTATA ATCGAATGAC GATTGCTATT AAAGATAATG	3360
25	GCATTGGTAT TCCTATCAAT AAAGTCGATA AGATATTCGA CCGATTCTAT CGTGTAGATA	3420
	AGGCACGTAC GCGTAAAATG GGTGGTACTG GATTAGGACT AGCCATTTTCG AAAGAGATTG	3480
	TGGAAGCGCA CAATGGTCGT ATTTGGGCAA ACAGTGTAGA AGGTCAAGGT ACATCTATCT	3540
30	TTATCACACT TCCATGTGAA GTCATTGAAG ACGGTGATTG GGATGAATAA TAAGGAGCAT	3600
	ATTAAATCTG TCATTTTAGC ACTACTCGTC TTGATGAGTG TCGTATTGAC ATATATGGTA	3660
	TGGAACTTTT CTCCTGATAT TGCAAATGTC GACAATACAG ATAGTAAGAA GAGTGAAACG	3720
35	TAACCTTTAA CGACACCTAT GACAGCCAAA ATGGATACAA CTATTACGCC ATTTTCAGATT	3780
	ATTCAATTCGA AAAATGATCA TCCAGAAGGA ACGATTGCGA CGGTATCTAA TGTGAATAAA	3840
	CTGACGAAAC CTTTGAAAAA TAAAGAAGTG AAGTCCGTGG AACATGTTTCG TCGTGATCAT	3900
40	AACTTGATGA TTCCTGATTT GAACAGTGAT TTTATATTAT TCGATTTTAC GTATGATTTA	3960
	CCGTTATCAA CATATCTTGG TCAAGTACTG AACATGAATG CGAAAGTACC AAATCATTTTC	4020
45	AATTTCAATC GTTTGGTCAT AGATCATGAT GCTGATGATA ATATCGTGCT TTATGCTATA	4080
	AGCAAAGATC GCCACGATTA CGTAAAATTA ACAACTACAA CGAAAAATGA TCATTTTTTA	4140
	GATGCATTAG CAGCAGTGAA AAAAGATATG CAACCATACA CAGATATCAT CACAAACAAA	4200
50	GATACAATTG ATCGTACGAC GCATGTTTTT GCACCAAGTA AACCTGAAAA GTTAAAAACA	4260
	TATCGCATGG TATTTAACAC GATTAGTGTT GAGAAAATGA ATGCTATACT ATTTGACGAT	4320
55		



EP 0 786 519 A2

	GCAAACTATA ACGATAAAAA TGAAAAATAT CATTATAAAA ACCTGTCCGA AGATGAAGCG	4440
	AGTTCCAGCA AAATGGAAGA AACGATTCCA GGAACCTTTG ATTTTATTAA TGGTCATGGT	4500
5	GGTTTCTTAA ACGAAGACTT TAGATTGTTT AGTACGAATA ATCAGTCAGG CGAGTTAACA	4560
	TATCaACGTT TCctTAATGG TTATCCAACG TTTAATAAAG AAGGTTCTAA TCAAATTCAA	4620
	GTCACTTGGG GTGAAAAAGG CGTCTTTGAC TATCGTCGTT CGTTATTACG CACCGACGTT	4680
10	GTTTTAAATA GTGAGGATAA TAAATCGTTG CCGAAATTAG AGTCTGTACG TTCAAGCTTA	4740
	GCGAACATA GTGATATTAA TTTTGAAAAA GTAACAAACA TCGCTATCGG TTACGAAATG	4800
	CAGGATAATT CAGATCATAA TCACATTGAA GTGCAGATTA ACAGTGAACG CGTACCGCGT	4860
15	TGGTATGTAG AATATGATGG CGAATGGTAT GTTTATAACG ATGGGaGGCT TGaATAAATG	4920
	AACTGGaAAC TGACAAAGAC ACTTTTCATT TTCGTGTTTA TTCTTGTCAA CATCGTGTTA	4980
20	GTATCGATTT ATGTTAATAA AGTCAATCGC TCACACATTA ATGAAGTCGA GAGTAACAAT	5040
	GAAGTTAATT TTCAGCAAGA AGAAATTAAA GTACCGACTA GTATATTGAA TAAATCAGTT	5100
	AAAGGTATAA AATTAGAGCA AATTACAGGG CGATCAAAAG ACTTTAGTTC TAAAGCTAAA	5160
25	GGCGATTCCG ATTTGACCAC ATCAGATGGT GGAAATTAT TGAATGCGAA CATTAGTCAA	5220
	TCGGTAAAGG TCAGTGACAA TAACTTAAAA GATTTGAAAG ATTATGTTAA CAAGCGCGTA	5280
	TTTAAAGGTG CTGAATATCA ATTAAGCGAG ATTAGTTCAG ATTCTGTAAA ATATGAACAA	5340
30	ACGTATGATG ATTTTCCGAT TTTAAATAAC AGTAAAGCGA TGTTAAACTT TAATATAGAA	5400
	GATAACAAAG CGACTAGTTA TAAACAATCA ATGATGGATG ACATTAAGCC CACAGATGGT	5460
	GCAGATAAGA AGCATCAAGT GATTGGTGTG AGAAAAGCAA TCGAGGCATT ATATTATAAT	5520
35	CGTTACTTGA AAAAAGGTGA TGAAGTCATT AATGCTAGAC TCGGTTACTA CTCAGTCGTG	5580
	AATGAAACGA ATGTTCAATT GTTACAACCA AACTGGGAAA TTAAAGTGAA GCATGACGGT	5640
40	AAGGATAAAA CGAATACTTA CTATGTCGAA GCGACAAATA ATAACCCTAA AATTATTAAT	5700
	CATTAATATG AATCGTAATA AGCTAGCATT GCAAGCTCAT CATATGTGAG AAGCGGTGCT	5760
	AGCTTTTTTG CTGGTACGGT TTATTATGGC TGATGTTTTT GCGTCTCAA CGTGCGCATT	5820
45	TATTCATATT TTAAGTAGAA CCGCATTGTA AAATTAGTGT AACTGTTATT TTAAAACTT	5880
	TAGTATTGT CTAATCATTG TTATAATAAT TAAGAAATTC ATTGCACGTG ATTATCAAAA	5940
	TTTAAATATA AGAAACCGGT CGATGAACTA AAGTTACATA ATAGGAAAGG TATACAAAAC	6000
50	AGCTAATATA CTGATAGTTT CTGTAGGGAA AATCGTATAT TTGCACTGAT GTATATTGCA	6060
	GTCATATAGA GAGATTGACT GTTTAAAGAG AAAGGATGAG CCGCTTGATA CGCATGAGTG	6120

55

	TAGTTGATGT TGGTTTGACT GGAAAGAAAA TGGAAGAATT GTTTAGTCAA ATTGACCGTA	6240
	ATATTCAAGA TTTAAATGGT ATTTTAGTAA CCCATGAACA TATTGATCAT ATTAAAGGAT	6300
5	TAGGTGTTTT GGCGCGTAAA TATCAATTGC CAATTTATGC GAATGAAAAA ACTTGGCAGG	6360
	CAATTGAAAA GAAAGATAGT CGCATCCCTA TGGATCAGAA ATTCATTTTT AATCCTTATG	6420
	AAACAAAATC TATTGCAGGT TTCGATGTTG AATCGTTTAA CGTGTCACAT GATGCAATAG	6480
10	ATCCGCAATT TTATATTTTC CATAATAACT ATAAGAAGTT TACGATTTTA ACGGATACGG	6540
	GTTACGTGTC TGATCGTATG AAAGGTATGA TACGTGGCAG CGATGCGTTT ATTTTTGAGA	6600
15	GTAATCATGA CGTCGATATG TTGAGAATGT GTCGTTATCC ATGGAAGACG AAACAACGTA	6660
	TTTTAGGCCA TATGGGTCAT GTATCTAATG AGGATGCGGC TCATGCAATG ACAGACGTGA	6720
	TTACAGGTAA CACGAAACGT ATTTACCTAT CGCATTTATC ACAAGACAAT AACATGAAAG	6780
20	ATTTGGCGCG TATGAGTGTT GGCCAAGTAT TGAACGAACA CGATATTGAT ACGGAAAAAG	6840
	AAGTATTGCT ATGTGATACG GATAAAGCTA TTCCAACGCC AATATATACA ATATAAATGA	6900
	GAGTCATCCG ATAAAGTTCC GCATTGCTGT GAGACGACTT TATCGGGTGC TTTTTTATGT	6960
25	TGTTGGTGGG AAATGGCTGT TGTGAGTTG AATCGGCTTG ATTGAAATGT GTAAAAAAT	7020
	TCGATATTAA ATGTAATTTA TAAATAATTT ACATAAAATC AATCATTTTA ATATAAGGAT	7080
	TATGATAATA TATTGGTGTA TGACAGTTAA TGGAGGGAAC GAAATGAAAG CTTTATTACT	7140
30	TAAAACAAGT GTATGGCTCG TTTTGCTTTT TAGTGTAATG GGATTATGGC AAGTCTCGAA	7200
	CGCGGCTGAG CAGCATAAC CAATGAAAGC ACATGCAGTA ACAACGATAG ACAAGCAAC	7260
	AACAGATAAG CAACAAGTAC CGCCAACAAA GGAAGCGGCT CATCATTTCTG GCAAAGAAGC	7320
35	GGCAACCAAC GTATCAGCAT CAGCGCAGGG AACAGCTGAT GATACAAACA GCAAAGTAAC	7380
	ATCGAACGCA CCATCTAACA AACCATCTAC AGTAGTTTCA ACAAAGTAA ACGAAACACG	7440
40	CGACGTAGAT ACACAACAAG CCTCAACACA AAAACCAACT CACACAGCAA CGTTCAAATT	7500
	ATCAAATGCT AAAACAGCAT CACTTTCACC ACGAATGTTT GCTGCTAATG CACCACAAAC	7560
	AACAACACAT AAAATATTAC ATACAAATGA TATCCATGGC CGACTAGCCG AAGAAAAAGG	7620
45	GCGTGTCATC GGTATGGCTA AATTAAAAAC AGTAAAAGAA CAAGAAAAGC CTGATTTAAT	7680
	GTTAGACGCA GGAGACGCCT TCCAAGGTTT ACCACTTTCA AACCAGTCTA AAGGTGAAGA	7740
	AATGGCTAAA GCAATGAATG CAGTAGGTTA TGATGCTATG GCAGTCGGTA ACCATGAATT	7800
50	TGACTTTGGA TACGATCAGT TGAAAAAGTT AGAGGGTATG TTAGACTTCC CGATGCTAAG	7860
	TACTAACGTT TATAAAGATG GAAAACGCGC GTTTAAGCCT TCAACGATTG TAACAAAAAA	7920

55

TGAAGGCATT AAAGGCGTTG AATTTAGAGA TCCATTACAA AGTGTGACAG CGGAAATGAT 8040  
 GCGTATTTAT AAAGACGTAG ATACATTTGT TGTATATCA CATTAGGAA TTGATCCTTC 8100  
 5 AACACAAGAA ACATGGCGTG GTGATTACTT AGTGAACAA TTAAGTCAA ATCCACAATT 8160  
 GAAGAAACGT ATTACAGTTA TTGATGGTCA TTCACATACA GTACTTCAA ATGGTCAAAT 8220  
 TTATAACAAT GATGCATTGG CACAAACAGG TACAGCACTT GCGAATATCG GTAAGATTAC 8280  
 10 ATTTAATTAT CGCAATGGAG AGGTATCGAA TATTAAACCG TCATTGATTA ATGTTAAAGA 8340  
 CGTTGAAAAT GTAACACCGA ACAAAGCATT AGCTGAACAA ATTAATCAAG CTGATCAAAC 8400  
 ATTTAGAGCA CAACTGCGAG AGGTAATTAT TCCAAACAAT ACCATTGATT TCAAAGGAGA 8460  
 15 AAGAGATGAC GTTAGAACGC GTGAAACAAA TTTAGGAAAC GCGATTGCAG ATGCTATGGA 8520  
 AGCGTATGGC GTTAAGAATT TCTCTAAAA GACTGACTTT GCCGTGACAA ATGGTGGAGG 8580  
 TATTCGTGCC TCTATCGCAA AAGGTAAGGT GACACGCTAT GATTTAATCT CAGTATTACC 8640  
 20 ATTTGGAAAT ACGATTGCGC AAATTGATGT AAAAGGTTCA GACGTCTGGA CGGCTTTCGA 8700  
 ACATAGTTTA GGCGCACCAA CAACACAAA GGACGGTAAG ACAGTGTTAA CAGCGAATGG 8760  
 CGGTTTACTA CATATCTCTG ATTCAATCCG TGTTTACTAT GATATAAATA AACCGTCTGG 8820  
 CAAACGAATT AATGCTATTC AAATTTTAAA TAAAGAGACA GGTAAGTTTG AAAATATTGA 8880  
 TTTAAACGT GTATATCAG TAACGATGAA TGACTTCACA GCATCAGGTG GCGACGGATA 8940  
 30 TAGTATGTTC GGTGGTCTTA GAGAAGAAGG TATTTTATTA GATCAAGTAC TAGCAAGTTA 9000  
 TTTAAAAACA GCTAACTTAG CTAAGTATGA TACGACAGAA CCACAACGTA TGTTATTAGG 9060  
 TAAACCAGCA GTAAGTGAAC AACCAGCTAA AGGACAACAA GGTAGCAAAG GTAGTAAGTC 9120  
 35 TGGTAAAGAT ACACAACCAA TTGGTGACGA CAAAGTGATG GATCCAGCGA AAAAACCAGC 9180  
 TCCAGGTAAA GTTGTATTGT TgtAGCGCAT AGAGGAACTG TTAGTAGCGG TACAGAAGGT 9240  
 TCTGGTCGCA CAATAGAAGG AGCTACTGTA TCAAGCAAGA GTGGGAAACA ATTGGCTAGA 9300  
 40 ATGTCAGTGC CTAAAGGTAG CGCGCATGAG AAACAGTTAT TTCATAATCA ACAGTCATTG 9360  
 ACGTAGCTAA GTAATGATAA ATAATCATAA ATAAAATTAC AGATATTGAC AAAAAATAGT 9420  
 45 AAATA 9425

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3886 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

	AGTTGTAATG TCACATTTCC AGAGTCTGAA ATTATCTTTA TCACGTTACA TTTACTAGGC	60
5	TCTAAAATGA CTGAACATAC AGCATCTTCA ATTACCTTTG AATACCATGA TTTATCGCAA	120
	AATATACATG AATTGATCAC TTGTGTTAGC CAAGAATTAG GCATTGATAT GTCAAAAGAC	180
	AACAAGTTAC ATACCAGTCT GATCACACAT ATCAAACCAG CTATACATCG TATTAAATAC	240
10	GATATGCTAC AACCTAATCC TTTGAGGCAA GAAGTTATGC GTCGCTATCC TCAAATCATT	300
	GAAGCCGTTA GCAAGCATAT TAGTCCAATT GAACAAGATG CTGCTATTTCG CTTCAACGAA	360
	GATGAATTAA CATAATTAC AATTCACCTC GCATCAAGTA TAGAGCGTGT TGCAACACAT	420
15	AAACAATCAA TGATTAAGGT TGTCTTACTA TGTGGTTCTG GTATAGGCAC GTCACAACCTT	480
	TTAAAATCAA AACTAAATCA CCTGTATCCT GaGTTnCACa TTTGGGAtGc CTATTcCATT	540
20	TaTcAATTGG aAGaAAGTCG ATTATTGCAA GATAACATTG ATTATGTCAT TTCAACAGTA	600
	CCTGTGAAA TATCAGCTGT ACCAGTTATT CATGTCGATC CATTTATCAA TCAACAATCT	660
	CGTCAAAAAT TGAATCAAAT TATCAATGAC TCAAGAGAAC AACGAGTCAT GAAAATGGCA	720
25	ACTGATGGCA AGTCACTCGC AGATTTATTG CCTGAACATC GCATCATTAT AAATAAACAA	780
	CCATTATCAA TTGAATCCGC AATTGCAGTG GCTGTGCAAC CTTTAATCAA TGATGGCATT	840
	GTCTATTCAA ATTATACAGC TGCAATTTTA AAACAATTTG AACAATTCGG GTCATATATG	900
30	GTCATTAGTC CACATATTGC ACTTATTCAC GCTGGTACTG ATTATGTACA GAATGGTGTA	960
	GGTTTCGCAC TAACATATTT CACTGAAGGG ATTATCTTTG GTAGTAAAGC TAACGATCCC	1020
	GTTACCTTG TAATTACATT AGCAACGGAC CACCCCAATG CACATTTAAA GGCATTGGGA	1080
35	CAGTTAAGCG AATGCTTAAG CAACGACTTA TATCGACAAG ATTTCTTAGA TGGGAATATT	1140
	TTTAAATTA AACAACACAT TGCTTTAACT ATGACAAAGG AGGCTTAATA ACGTGTCAAT	1200
	AGACATTTTG TCAACAACAC GCATCATTGT AAAAGAACAA GTAAATGATT GGAATGAAGC	1260
40	TATAACTATA GCTTCTCAGC CATTACTACA AGAACAAATT ATTGAACAAG GCTATGTTCA	1320
	AGCAATGATT GATAGCGTTA ATGAACCTTG ACCTTATATC GTTATCGCAC CTGAAATTGC	1380
45	AATTGCACAT GCAAGACCGA ACAATGACGT ACATCAAGTT GGTTTAAGTC TATTAAAGTT	1440
	GAATCAACAT GTGGCATTTC GTGATGAAGA TCACTACGCA TCTCTCATTT TTGTATTGAG	1500
	TGCCATCGAC AATCATTAC ACTTATCTGT ATTACAAAAT TTAGCAACCG TACTGGGCGA	1560
50	TAACCAAACA GTCCAGCAAC TATTAAGTGC AACAAATGCA CAAGACATTA AAAACATTTT	1620
	AAAGGAGCAT GATTAATATG AAAATTTTAG TAGTATGTGG CCACGGTTTA GGAAGTAGTT	1680

55

	AAGTTGAACA TAGTGACATT ATGACAGCAA GTCCAGAGAT GGCTGACTTG TTTATTTGTG	1800
	GTAGAGATTT AGCTGAAAAT GCCGAACGTC TAGGGGATGT CTTAGTTCCTT GATAATATTT	1860
5	TAGATAAAGC TGAATTACAA CAAAAGCTCT CAGAAAAATT ACAACAACCTT AACATGATTT	1920
	AAAGGAGGTA CGACCTATGC AAGCAATCCT TAATTTTATA GTCGATATTT TAAGTCAACC	1980
	AGCCATTCTT GTTGCACTGA TTGCCTTTAT AGGTTTAATC GTTCAGAAAA AACCTGCCGC	2040
10	AACGATCACT TCAGGAACCA TTAAAACGAT ATTAGGCTTC TTAATTTTAA GTGCAGGTGC	2100
	TGATGTCGTC GTTCGATCTC TTGAACCATT CGGCAAAATA TTCCAACACG CATTTGGTGT	2160
	GCAAGGTATC GTACCTAACA ACGAAGCTAT CGTCTCACTA GCCTTAAAAG ATTTTGGAAC	2220
15	AACAGCTGCA CTCATCATGG TCTGTGGCAT GATTGTTAAT ATTTTAATTG CCCGCTTCAC	2280
	TAATTTAAAA TATATCTTTT TAACAGGTCA TCATACATTT TACATGGCTG CGTTTTTAGC	2340
20	AATCATTTTA ACAGTCAGTC ATATTAAAGG CTGGCTAACG ATTGTTATCG GCGCACTCGT	2400
	ATTAGGATTA ATCATGGCAG TATTACCTGC ATTACTCCAA CCTACGATGC GAAAAATTAC	2460
	AGGGAATGAC CAAGTAGCTT TAGGTCATTT TGGCTCAATC AGTTACTTTG CCGCAGTGCT	2520
25	GTAGGTCAAT TATTCAAAGG TAAGTCTAAA TCAACGGAAG AGATTAAATT TCCAAAAGGC	2580
	TTAAGTTTCT TACGAGAAAG TACAATTAGT ATCTCGATTA CGATGGCATT ACTTTACTTC	2640
	ATCGCATGCT TATTTGCGGG CGTTAGTTAT GTACACGAAT CTATTAGTGA TGGTCAAAAC	2700
30	TTTATTGTCT TTTCATTAAT TCAAGGTGTG ACATTTGCTG CTGGTGTATT TATTATTTTA	2760
	ACGGGCGTTC GTTTAATCTT AGCTGAAATC GTCCCAGCAT TTAAAGGAAT TTCTGAAAAG	2820
	CTTGTACCAA ATTCTAAACC TGCATTAGAC TGCCCTATTG TGTTCCTTA TGCACAAAAT	2880
35	GCAGTATTAA TTGGATTCTT TGTCAAGCTT ATTACAGGTG TCATCGGTAT GTTTATCTTA	2940
	TTCTTATTTG GTGGCGTCGT CATTTTACCT GGCCTAGTTG CACACTTCTT CTTAGGTGCA	3000
	ACGGCTGCTG TATTCGGTAA TGCAAGAGGC GGTATTAAAG GTGCTATTGC TGGCGCCGCT	3060
40	CTAAATGGTA TCCTAATCAC GTTTTTACCA TTATTATTCT TGCCATTTTT AGGCGAATTA	3120
	GGTGGTGCTG CAACAACATT CTCAGATACA GACTTTTATG CTGTCGGTAT CGTGTTCCGT	3180
45	AACGCAGTAA AATATATGGG ATTATTTGGT GCGATTCTAT TTATTATTAT CGTAGGTGCG	3240
	ACAACAATTT TATTAAAAGG CCGTCAAAAA GAACAGCAAT AGTGTTAACG TAGAAATATA	3300
	AAACACCGTC ACATATTGAG TGAATGCCCC TTTTATCAAG AGGAAAGCCA CTTACTTATG	3360
50	GACGGTGTTT TGTATTATAT TAAATGATAC TTAGCCATAC TATCGACAGC TGCTAAAATT	3420
	GCTTCTTCTT GTGTCGCAAT CGGTTCCCAA CCAAGTAATG TTTTTGCACG TTCGTTACTT	3480
55		

CCTAGACTCA AAATAAAGTC TGGTAATTTT TTAGTAGAAA CTTTTTGAGC TATTTTCAGGT 3600  
 CTCTTTTCTT TAATTAATTT TGCAATTTCC AACAAATTAA TTGTCCATC AGCCGTCGCA 3660  
 5 ATAAATCGCT TGCCATTAGC TTGTTCAATT GTCAATTGCCA AAATGTGCAG TTCAGCTACG 3720  
 TCTCTCACAT CAACAACATT TAACGGAATT TCGGTACAC GTTTCATTGA ACCATTCAAT 3780  
 AAATTTTCTA ATAAATGAAA GCTTCCTGAA ACGTGTGCAT CTAATGATGG CCCAAAAATT 3840  
 10 GCAACTGGAT TGATTGTGGC AAATTCTACT GTTGTATTTT CATTCT 3886

## (2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 4879 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GTCATCTATC AAAAATTTGG TATACAGACC GACAATTATT AATTAATAAT TTAATTTCCC 60  
 25 AGGCAATACC AGTGATTAAA TATCCACAAA TACAACATAA AGAACAACCA TTAGAATCTA 120  
 TTTCACAAC TATAATGTCT AAGATGACAT CTAATCAATA GTGTTTAAAT TTCTCAGTGG 180  
 CTGTGAATGA GGTTTAAAAG TACTATAAAA CGTAAACTTT GATACTTTAA AATACGCAAA 240  
 30 AAACGGTAAA CCCTAATCA TATTATAGAG TTTACCGTTT TATTTTTTAA CTTGCATCAT 300  
 AGTTATATTA ACATTATTGT TGGTAGTTTG GATCAGTAAC CATTGCTTGT CCAGTATAAT 360  
 CAACCGTTAC AATTGAATAT TTTCCATTG CATTGCGTC TTTAAACTA AACACATACT 420  
 35 TATAGTTGCC ATTATGTTCT TCAATAGAAT AATCATTATA CACTTTATTA TTAACACAA 480  
 ATTTATTGTC TTCATTATTA GCCGCATTTA AAGCTGTTTG GAAATTGGC AATTGCTGTA 540  
 AAGCTTGATT TTTATTTCCA TTAAACGGAT AAATTTGACG TGCAACCGGC GCGGCATTTT 600  
 40 GACCATAATA TGGTGCAACG TAACTTGATT TTTGATTATT ATTCGCTTGG TTATTACTTG 660  
 ATTGGTTATT ATTTGTTTGG TTTTGGTCAT TGTTTGTGC ATTTGAATTA GATTGTGCT 720  
 45 GGTATCGTT TGCATTATTA TCTTTATTAT CTTTGTTTAC GTCTTTACTA TCATCTTTAT 780  
 TATCTTTCTT ATCTTTAGAT GAATCATTTG TTTTTTATC TTGTTGTTCA GTTTTCGCTT 840  
 TATCATCTTT TTCTTTATTA CCGTCTTTTT GTTGGTCACT ATCTTGACCA CATGCAGCTA 900  
 50 AAAATAATGA TAATGCTAGT AACCTGTAA CTAATCTTTT CACATATATC TCCTCCTATA 960  
 ATTCGATATT CATTGAATAA TCTTGAAATA CATATCTACC ATGTGTATCT TTTCATGGCT 1020

55

EP 0 786 519 A2

	TAAGGTTCTT	TTTATTATAC	CCTAATTTTT	GTTTCATTATT	ATTTAATTTT	TGTGAATTTT	1140
	ATGcTTkCTA	TAAATTTAAT	TATTTTACTT	TAACAATTCA	TTACGCATT	AGCATTTCAA	1200
5	GGTATACACA	ATATTTATTA	CTATGATTTC	ATTTTATCTG	CTGCAAAAAC	AATCATTATA	1260
	ACTCTTTTTC	CATAATTAAA	TCTGTATCCG	TTACATCACC	TGTTTGAAAA	TGATGTTTAC	1320
	CAACCACTTT	AAATCCATGA	CGTTTATAAA	ATGCTTGAGC	ACGAGGATTA	TGCTCCCAAA	1380
10	CTCCTAGCCA	AATTTTATGT	TTATTATGTT	CTTGAGCAAT	TTTTTCGGCC	AATTCTATCA	1440
	ATTGTGAACC	TCTTCCGCCA	CCTTGAAAGT	CTTTCAAAAA	ATATATGCGC	TGCACTTCTA	1500
	AATAGGTCTC	CCCCATTTCT	TCAGTTTGAG	CACTATTAAT	ATTCATCTTT	ATATAACCAA	1560
15	CATTTCGCACC	ATCTTCTTGa	TAAAAATAAT	GAAATGAATC	TACATGGTTA	ATCTCTTGTC	1620
	TAAATTTCCTC	TACAGTATAA	TTGTCTTTAA	AAAATGATC	AAAATCTTTG	TCATCATAGT	1680
20	AAGAACCAAA	CGTGCATAA	AATGTTCTAG	TTGCTAATTC	AACTAATTCA	CTAGCATTTT	1740
	GTTCTGAAAT	TTCTTTGATT	ATCCCAGCCA	TATAAATCCT	CCAATAAACA	GTGATCGAAT	1800
	CAAAATATTA	CTTATGTTAT	TTTTCAGCCA	AACTATTTA	AAAATACATT	AACACAAATC	1860
25	AATTACAAAT	TGTATTGATT	GTGTGTAACA	TCAATAAATG	ATACATTTAT	TCCAGTAAAA	1920
	TGGCCGTATT	TTCAAAAGAG	AAAAAGAGAG	GATGTATCGT	TGTGATAGAA	ACATTTAAAG	1980
	CGTTTGAAT	TGATAAAGAT	GAGAGTGGTA	AAGTGACACC	AACTTTCAAA	CAATTATCGC	2040
30	CTACTGATTT	ACCTAAAGGA	GATGTGCTGA	TTAAAGTACA	TTACTCTGGT	ATAAATTATA	2100
	AAGATGCTTT	AGCGACTCAA	GATCATAATG	CAGTCGTAAA	ATCGTATCCT	ATGATTCCAG	2160
	GAATAGATTT	AGCTGGAACA	ATTGTTGAAT	cCGAAGCACC	AGGCTTTGAA	AAAGGAGAAC	2220
35	AAGTAATTGT	AACGAGTTAT	GACCTAGGTG	TCAGCCATTA	TGGCGGTTTT	AGTGAATATG	2280
	CGCGTGTAAA	ATCAGAATGG	ATTATCAAGC	TTCCTGATAC	TTTAACATTA	GAAGAATCAA	2340
40	TGATATATGG	CACAGCTGGT	TATACTGCCG	GTTTAGCAAT	TGAAAGACTT	GAAAAAGTTG	2400
	GAATGAATAT	TGAAGATGGT	CCTGTACTCG	TTCGCGGTGC	TTCAGGTGGT	GTCGGTACTT	2460
	TAGCAGTACT	CATGCTTAAT	GAACTTGGTT	ATAAAGTTAT	CGCAAGTACA	GGTAAACAAG	2520
45	ATGTTAGCGA	TCAATTACTT	GAACTTGGTG	CCAAAGAAGT	TATCGATCGA	CTTCCTGTTG	2580
	AAGATGATCA	TAAAAAGCCA	CTCGCATCAT	CAACTTGGCA	AGCTTGTTGA	GACCCTGTTG	2640
	GTGGCGAAGG	TATTAATTAT	GTTACAAAGC	GTTTAAATCA	TAGTGGGTCA	ATTACAGTTA	2700
50	TTGGTATGAC	TGCCGGTAAT	ACTTATACTA	ATTCTGTATT	CCCTCACATT	TTAAGAGGTG	2760
	TAAACATTTT	AGGAATTGAC	TCGGTATTTA	CTGCTATGAA	ATTAAGACAG	CGCGTTTGCC	2820
55							

	TTGATGAACT TCCAGAACAA CTTAACAAAG TAATTAAACA TGAAAATAAA GGGCGCATTG	2940
	TTATCGATTT CGGTGTAGAT AAATAGTATT CATGAAAAAG ACATCCCGTT ATGCGAGATG	3000
5	TCTTTTTTAA TTTAGTATTT GATATACATA CCGCCTGAAT CTGGTTCGGT AGGTATAAAT	3060
	CCAAATTTTG TATATAATTT ATCCGCTGGG TAGTCTGCAA TCAGACTAAC GTATGTACTC	3120
	TCAACAGCCA CACCTTTAAT ATATTGCATA ATATGCTCCA TAATTAGACT GCCGTAACCT	3180
10	TGACCTTGGT AACTTTTCAA AACTGCAATA TCAACAATTT GAAAAACAGT TCCGCCATCG	3240
	CCAATCACTC TACCCATACC AATTAACCGA TCTTTATCAT ACAAGGTTAC TGTAATAAAG	3300
	GCATTAGGTA ATCCTTTTTT aGCTGTTCGC GCGTCTTTGG ACTCATACCT GCGTTAATCC	3360
15	TTAATGCGCA ATAATCCTCG CAAGTCGGAA TATCATATGT CACTTTAACC ATTATTTACC	3420
	CCACTTTTCA TCACACAATA TATCAACCTA GTATAAATGT TTATTTACAA TAGTCTTATT	3480
20	CGCTTCTTTA AACACTTCAT GATGACTTGA AACATAACCC TCTGCATTCTG CATCTGGTTG	3540
	GATATATGTT TTAGCAAGGT TCGCTGCATT TGCACCATCA CTAAATGCAC TTGCAATTAG	3600
	ATGTGATTTT GCATCATGAT AAACAATATC TCCACACGCA TAGATACCAG GTATACTAGT	3660
25	TGTCGTATTA CCAAATCCTT TAACACGACA ATCATCATGC ATATCTAGCT TTGAAGATGT	3720
	TtCACTCAAT AATGTATTAC AACGATCAAA CCCATGACTA ATAATGACAT CGTCAAATTT	3780
	AACTGTATGC CTATCGCCAC TTTCAACATG TTCCAAAACA ACTTCACTTA TATGCGTTTC	3840
30	ATCATCATTG CCGACCAAGT ATTAAATACG TGTTTTTGGG CATAGTTTCA CATTTAAATC	3900
	TGTCACCAAC GTTTTTCATCG CTTTCATGACC ACTTACATCT TCTTTTCGAT AAACAACGTG	3960
	CACGCTTTTA GCAATCTTGG CAATATCATG CGCCCAATCT AATGCTGTAT TTCCTCCACC	4020
35	TGATATTAAT ACATCTTTAT CTTTGAAACG TCTGTAACCT TGTACAACAT AATGTAAATT	4080
	AGTTtAATTGA TATCTCTCTA CACCTTTAAT ATCTAATTGT TTTGGATTAA TAATACCGC	4140
	ACCAATTGCA ATGATAACTG CTTTCGATGT ATATATTTCT CCCGCTTCTG TTTCAACTTC	4200
40	GAAATGACGT TCTGCCTTTT TCCTAATATC TACCACACGT TCATTCAAAT GAACTTCCGG	4260
	TTTAAAATAT AATCCTTGCT TAATTGTATC TTTTAAATTT TCATGACAAG GTTTTGGCGC	4320
45	AATGCCGCCA ATATCCCAA TAATTTTTTC AGGGTAAATT CTCATCTTAC CCCCTAATTC	4380
	AGATTGAACA TCTATCAATC TTACAGACAT ATCTCGCAAT CCAGCATAAA AGCTTGcata	4440
	CAAACCAGAC GGACCGCCAC CAATGATTGT AACATCTTTC ATTATGTGCC TCCTATGACT	4500
50	CTCTATATTC ATTTCTTTCA TTAACGTGCT CAAATTGATA ATTATTATCA TTTAAAGCCA	4560
	TTATACTATT AATATTTATA TTGTTAAAT AAATCGCATA GTTAGCCATG AATTATCAAT	4620

55



GAAAGATGTG TATATTTTTT AGTTCTAGTT ATATTATTTT TTAAAAGACT CATCACGTGG 4740  
 TTCTTTAAGA ATTGCTTGTC TTAAAAGGAA AAATAGCAAC AATAAACCTG CAAGCATACC 4800  
 5 TGTGTGCCCA ATACCTGCAA AGCCTGChAA TGCTTCTGGA GAGTATGATT TACCAGTGAC 4860  
 TTGGAAGAAT CCTTTTGTC 4879

(2) INFORMATION FOR SEQ ID NO: 90:

10

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1560 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

20 ATAATGTCTT AGaTTGATTG GGAGTTTTTTT TAAITTTTTTT GAAATTAAAT TAATCTGTAS 60  
 yTAATAAAAA ATTTGAATAA CTGACACaYT TTTTGTATCA TAGCTAyATA CTTTGTGAAT 120  
 TAATTCACAT TATAATAAGA GTGAAGATAA GAGTATTATA AATnATCTTT AAATAAATAT 180  
 25 ATGTGAAGTA AAAATTACAC GTTAGCATAT CGATTATGgT CATTTckTTT AACATATTAA 240  
 CTgGGGaACG TTAAAAGTTA ACGGkTGATA TCyAACTAAA AACAAGGTCA CAGTAGTATG 300  
 TTTTAATCTG GCGTCTATTA CAAATAAAAA TTACATCTAT AATTATTCGT TTTCTTTTTT 360  
 30 GAAAGTAATA GCCAATTAAT ATCATACATA CTGGAGTGAC TATAAGGAGG ACATTATTAT 420  
 GAGAGCAGCA GTTGTAAACGA AAGATCACAA AGTAAGTATT GAGGACAAAA AGTTAAGAGC 480  
 TTTAAAACCT GGTGAAGCGT TGGTACAAAC GGAATATTGT GCGGTTTGTC ATACCGATTT 540  
 35 ACATGTTAAG AATGCTGATT TTGGTGATGT TACAGGCGTT ACTTTAGGTC ATGAAGGTAT 600  
 TGGTfAAAGTC ATCGAAGTTG CGGAAGATGT AGAATCATTa AAAATTGGAG ACCGTGTGTC 660  
 40 TATCGCTTGG ATGTTcGAAA GCTGTGGAAG ATGTGAATAT TGTACAACAG GTCGTGAAAC 720  
 ACTTTGCCGT AGTGTGAAAA ATGCTGGTTA TACAGTAGAT GGTGCAATGG CTGAACAAGT 780  
 TATTGTTACT GCAGACTATG CTGTGAAAGT ACCTGAAAAA TTAGATCCAG CAGCAGCGTC 840  
 45 TTCTATTACA TGCGCAGGTG TGACAACCTTA TAAAGCTGTA AAAGTAAGTA ATGTAAAACC 900  
 TGGACAATGG TTAGGTGTTT TTGGTATAGG TGGTTTtagGT AACCTAGCTT TACAATATGC 960  
 TAAAAACGTT ATGGGGGCTA AAATTGTTGC CTTcGACATC AATGATGATA AATTAGCATT 1020  
 50 CGCGAAAGAA TTAGGTGCTG ATGCTATTAT TAATTCTAAA GATGTTGATC CAGTTGCAGA 1080  
 AGTTATGAAA TTAAGTATA ACAAAGGATT AGATGCAACA GTGGTAACTT CAGTTGCTAA 1140

55

TTTACCTGTT GATAAAATGA ACTTAGATAT CCCAAGATcA GTGCTTGATG GTATTGAAGT 1260  
 AGTAGGTTCA CTTGTTGGTA CAAGACAAGA CTTACGTGAA GCGTTTGAAT TTGCTGCTGA 1320  
 5 AAATAAAGTA ACACCTAAAG TTCAATTAAG AAAATTAGAA GAAATCAATG ATATTTTGA 1380  
 AGAAATGGAA AATGGTACTA TAACTGGTAG AATGGTTATT AAATTTTAAA AATATCAACT 1440  
 GACTATATAG ATAAAGAAGG TAGTGCTCTG AACACTATCA TTATTAATCA AACCCCGAGG 1500  
 10 TTTTCCTGAA AAGATAGTGG nAAATCCCCG TGTTTTTTGG GTTTGAGGnG GTTGnTGTA 1560

(2) INFORMATION FOR SEQ ID NO: 91:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11014 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

GTCCTGTnGC TGCAATGAAT ACGCCTAAAA ATCCAGGGAT GTAATGGATA CTTTGTGGTA 60  
 25 GTACTAATGA TAGAAATGAT AAAAATGAAA TCACAAAGGC TACGCTCGCA AAAGCTTGAC 120  
 ATGTACGCTT ATCGCCATAA TCTAACCTTG TACGTATATG TAATAAATAC TGTAATCCGA 180  
 TACTTAAATA CATAATTGCC ACGCATAAGA AGAATGGGAA GAATGTCTTT TCAAAGTCCG 240  
 30 GATATAGGCT GTTAGATAGG AAGACCATGA TGAACATATT AAACATCATA AACGAGACGT 300  
 CTTTGAATGT AACTTGACCA AATCGATTTG TAAAAAATGT TTGATGAGAC CACATTAACC 360  
 ATAAGAACAA ACTCATGACG ATGTATTTGA AAAATAAATC AGCTGAAATG GAACCGTTTT 420  
 35 GTGTTGTTAA AATCACATGT GCAATTTTTT GAATGGCATA GACGAAAATT AAATCAAAGA 480  
 ACAATTCATG GAATCCTGCA CGCTTTTCAG CTAAATGTTT TGGTGTAAAT GCATTAACCA 540  
 TAAAATTTTA ACTCCTTTAA GATGTGTAAT TAATTTACTA AGTATACTAT TTATTTTTC 600  
 40 TAGTGAATAG GGGCAGATTT GGCATGAAG TGAAGGAGA GGTGACTGCA AGGTAATTGC 660  
 GGAATTAACA ATCATCAGCG ATTTAATATT TGAATGGAGA CGTCATGGTA ATAAAAAATT 720  
 45 GATGAGAAAT TGATGGTGAA ACCAGCTGTG AATAsCGaTG cAATGATrsA TAGaATTTAA 780  
 TTAGAGTCAT TACGCGaAAT GATTAATGAT AATTTGTGGT AAATCAAAGC aTAATTTTGT 840  
 ACTATAGATG AGGATGATAG AGCATATTTA AGAGGGTGAA ATGTTAAAGT GAAACCGTTT 900  
 50 ACGTTTCCGA TTGCCCAAAC AAATTACATC ATTGTATAAT ATGATTTGTT AAATGCATAA 960  
 CAAGAATGAA AATGTAACAT ACGTAGCAAT TGGTTTCATA AATTGGATGT TAGTGCGTA 1020

55

	TGACGAGAGT	CGTATTAGCA	GCAGCATACA	GGACACCTAT	TGGCGTTTTT	GGAGGTGCGT	1140
	TTAAAGACGT	GCCAGCCTAT	GATTTAGGTG	CGACTTTAAT	AGAACATATT	ATTAAAGAGA	1200
5	CGGGTTTGAA	TCCAAGTGAG	ATTGATGAAG	TTATCATCGG	TAACGTACTA	CAAGCAGGAC	1260
	AAGGACAAAA	TCCAGCACGA	ATTGCTGCTA	TGAAAGGTGG	CTTGCCAGAm	ACAGTACCTG	1320
	CATTTACGGT	GaATAAAGTA	TGTGGTTCTG	GGTTAAAGTC	GATTCAATTA	GCATATCAAT	1380
10	CTATTGTGAC	TGGTGAAAAT	GACATCGTGC	TAGCTGGCGG	TATGGAGAAT	ATGTCTCAAT	1440
	CACCAATGCT	TGTCAACAAC	AGTCGCTTTG	GTTTTAAAT	GGGACATCAA	TCAATGGTTG	1500
15	ATAGCATGGT	ATATGATGGT	TTAACAGATG	TATTTAATCA	ATATCATATG	GGTATTACTG	1560
	CTGAAAATTT	AGTAGAGCAA	TATGGTATTT	CAAGAGAAGA	ACAAGATACA	TTTGCTGTAA	1620
	ACTCACAACA	AAAAGCAGTA	CGTGCACAGC	AAAATGGTGA	ATTTGATAGT	GAAATAGTTC	1680
20	CAGTATCGAT	TCCTCAACGT	AAAGGTGAAC	CAATCGTAGT	CACTAAGGAT	GAAGGTGTAC	1740
	GTGAAAATGT	ATCAGTCGAA	AAATTAAGTC	GATTAAGACC	AGCTTTCAAA	AAAGACGGTA	1800
	CAGTTACAGC	AGGTAATGCA	TCAGGAATCA	ATGATGGTGC	TGCGATGATG	TTAGTCATGT	1860
25	CAGAAGACAA	AGCTAAAGAA	TTAAATATCG	AACCATTGGC	AGTGCTTGAT	GGCTTTGGAA	1920
	GTCATGGTGT	AGATCCTTCT	ATTATGGGTA	TTGCACCAGT	TGGCGCTGTA	GAAAAGGCTT	1980
	TGAAACGTAG	TAAAAAGAA	TTAAGCGATA	TTGATGTATT	TGAATTAAAT	GAAGCATTTG	2040
30	CAGCACATC	ATTAGCTGTT	GATCgTGAAT	TAAATTACC	TCCTGAAAAG	GTGAATGTTA	2100
	AAGGTGGCGC	TATTGCATTA	GGACATCCTA	TTGGTGCATC	TGGTGCTAGA	GTATTAGTGA	2160
	CATTATTGCA	TCAACTGAAT	GATGAAGTTG	AAACTGGTTT	AACATCATTG	TGTATTGGTG	2220
35	GCGGTChAAC	TATCGCTGCA	GTTGTATCAA	AGTATAAATA	ATAAGAAAAC	AGGTTATCAC	2280
	AACAStATTA	ATtACATGTT	GGCATAACCT	GTTTTTATTT	GTTTATGGAT	TTATTGGGTA	2340
	ATATTAGTCA	TTTGATGGTT	TAATTGCAAA	TGCTCTAACA	GGGAACCCAG	GTGCATCTTT	2400
40	TGGTTTAGGG	CTGATAGCGT	AAATGATGGC	GCCACGAGTT	GGTAATTGAT	CTAAATTAGT	2460
	TAATAACTCG	ACTTGGTATT	TATCCTGACC	AAGAATATAA	CGTTCGCCAA	CTAAATCACC	2520
45	ATTTTTTACA	ACGTCCACAG	ATGCATCGGT	ATCGAATGTT	TCATGACCAA	CAGCTTCAAC	2580
	ACGACGTTCT	TCAATTAAGT	ACTTCAAAGC	ATCTAATCCC	CAACCCGGTG	CATGTTGTTG	2640
	TCCGTTTCGCA	TCTTTGTTTT	CAAACTTTTT	AATATTAGGC	CAACGTTTTG	ACCAATCGGT	2700
50	ACGAAGTGCA	ACAAAAGTGC	CAGGTTCAAT	AGTACCATGC	TCTTTTTCCC	ATGCTTCTAT	2760
	ATGCGCACGT	GTTACGATGA	AATCATTGTT	GTTTCGCTACT	TCTGTTGAAA	AGTCTAATAC	2820

55

	AAAGTGAATT GGTGCATCAA TGTGAGTACC ATATTGCGTT ACAATATTCC AACGTTGCAC	2940
	ATAGAAACCA TGATCTTTAA CCGTGAATAA AGTTGAAACT TCGCCTTTTT CAAACTCACT	3000
5	AAAACGTGGT ATTTCCGGAT CAAATGTATG CGTTAAATCA ACCCAAGTTG CTTGTTTTAA	3060
	AGTATTTAAT TGTTGCCATA AAGGATATTG TGTACATAAAA TCACCCGTTT TTAGTTTATT	3120
	ATATGATAAA TGCTGCGATT ATTCTTGGCG TTTAGCTTTA ACAGCATTCA CAAGCACAGT	3180
10	CAATGCATCT TTAACCTTCTT CTTCTTTTCG CGTTTTTAAA CCACAGTCAG GGTTTACCCA	3240
	GAATAATGAG CCGTGCATTT GTTGTAGTGA ACGATTGATT GCTGTAGTAA TTTCTTCTTT	3300
15	TGTTGGAATA CGTGGACTAT GAATATCATA TACACCTAGA CCAATACCTA AATCATAATT	3360
	AATATCTTCA AAGTCTTTAA TTAAATCACC ATGGCTACGA GATGTTTCAA TTGAAATAAC	3420
	ATCAGCATCT AAGTCATGAA TAGCATGAAT GATTTGACCG AATTGAGAAT AACACATATG	3480
20	TGTATGGATT TGAGTTTCAT CACGAACTGA AGACGTTGCA AGTTTAAATG ATAAAACAGC	3540
	ATCTTTAAGA TATTGTTTCG GATATTCAGA GCGTAATGGT AAGCCTTCAC GTAATGCAGG	3600
	TTCGTCAACT TGGATAACTT TGATTCTGCG AGCTTCAAGT GCTAATACTT CTTGCTTGAT	3660
25	TGCTAAAGCA ATTTGATCTT GAACGACTTT ACGTGGTAAA TCAACACGTT CAAATGACCA	3720
	GTTTAGAATT GTTACAGGTC CAGTTAACAT ACCTTTAACT GGTTTATCTG TTAAGCTTTG	3780
	TGCATAAACT GTTTCATCAA CAGTTAAAGG CGCTGTCCAT TTTACATCAC CATAAATGAT	3840
30	TGGTGGTTTT ACGGCACGTG AACCATATGA TTGCACCCAA CCGAATTTAG TTAAGTAAAG	3900
	ACCTTGTAAT TTTTCTCCGA AGAATTCAAC CATGTCATTA CGTTCAAATT CACCGTGAAC	3960
	TAATACATCT AAGCCAATGT CTTCTTGAAT TTTAATCCAT CGAGCAATTT CATTTTTTAA	4020
35	GAATGTTTCA TATGCTTCGT CTGTAATGCG TTTGTTCTTC CAATCTGCAC GGTATTTTCG	4080
	AACCTTCTCGG CTCTGTGGGA ATGATCCAAT AGTTGTTGTT GGTAAATCCG GTAAGTTCAA	4140
40	ACGTTTTTGT TGTGTTTCAA TACGTTGCGC GAATGGTGAT TGTCTTGAAG TACGCACGCT	4200
	TTCGAAATCA TAATCTAAGT TTTTGAATGA TTGATTTTGG AAACGCTCAT AACGTGCTTT	4260
	TAATTTATCA TATTTAACAC TATCGTTTTG ATTAAATAGG CGACGCAATG CATCTAATTC	4320
45	GTCTAATTTT TCAGTTGCAA AGCTTAAGCC TTCGCCAACA CTTGTATCTA ATGTTTCATC	4380
	ATCTAAAGAT ACTGGAACAT GTAATAATGA AGATGATGGT TGAATGACAA GTTCATTAGT	4440
	GTGTGCTAAC AATTTATCGA TTAAGACTTT TTTAGCTTCA ATGTCACTTG CCCATACATT	4500
50	ACGACCATCA ATAATTCAG CGTATAATGT TTTTGATTTA TCAAAATCTC CAGCTTCAAT	4560
	TTGTTTAAGG TTATAGCCAT TATCATGGAC AAAGTCTAAA CCTATACCAC CAACAGGTAA	4620

55

	AACACCAGCT	TTTTCGAAAT	AGTCATAAGC	TTCACGTGTA	ATATTTTCAT	AGCTTTTCGCT	4740
	GTCGTCTGTA	ACTAAGATTG	GCTCATCAAC	TTGAATGTAC	TCAGCACCTG	CATCAATTAA	4800
5	TGATTCAAAC	ACTTCTTTAT	AAAGTGGTAA	TAACGTTTAA	ACTTTTCTT	CAAAAGTTTG	4860
	GTGACCGCCT	TTTGATAATT	TAACAAAAGT	AATCGGACCA	ACAATGACAG	GGTGAGCGTT	4920
	AACGTTTAAA	GATTGGGCAT	ATTTAAAGCG	ATCTAATAAT	ACATTGCGAC	TCACCTTAGG	4980
10	CTCAACATTG	TCCCATTGAG	GTACGATGTA	ATGATAGTTA	GTGTTAAACC	ATTTTATAAG	5040
	TGCACTTGCA	ACATGGTCTT	TATTACCGCG	AGCAATATCA	AATAATAAAT	CATCATCAAT	5100
	AGTTCTTCCT	TGAAACGTT	CAGGGATGAT	GTTGAATAAT	AATGACGTAT	CTAATATATG	5160
15	GTCATATAAA	GAGAAATCAC	CAACTGGGAT	GCTATCTAAG	TGATAGTACT	TTTGTAATAA	5220
	TAAATTTCT	TTATGTAGAT	CAGTTAATGT	TTGATCTAAT	TCTTCTTTAG	AAATCTTCTT	5280
20	TGCCCAATAA	CTTTCGATGG	CTTTTTTCCA	TTCTCTTTT	CTACCTAATC	TTGGGAATCC	5340
	TAAGTTTGAT	GTTTAAATTG	TTGTCATAAT	ATTGCCTCCT	TGTGAGCAGT	AATAGATTTT	5400
	GAGTATGCTG	CAAGTTCTAA	TGAATCTTCG	ACATTTTGAA	ACGGTGTGAT	AATGTATAAA	5460
25	CCATTAAAAT	ATTCATGAAC	AGTATCGATT	AAATCCTTTG	AAAGCTTAAG	ACTTAGTTCT	5520
	CGTGTTTTGG	CTTTATCATC	TTTAACTGCT	TCAAATGTG	GTAAATTTT	ATCTGACATC	5580
	TTGATTCCTG	GCACTTCATT	ATGCAAAAAG	AGTGCGTTTT	TGTAACCTGC	GATAGGCATA	5640
30	ATGCCTATGA	AAAATGGTTT	GTTCAAGTGC	TTAGTGGCAT	GGTAAATTTT	AATGATTTTC	5700
	TCTTTGCTGT	ACACGGGTTG	TGTTATAAAA	TAAGACATTC	CGCTTTCTAT	CTTTTTCTCT	5760
	AATCTTTTGA	CGGCACCATA	TAATTTACGA	ACATTAGGGT	TAAAGGCGCC	AgcGATGTTG	5820
35	AAGTGTGTAC	GTTTCTTCAG	CGCATCACCG	TCAGTGTTAA	TACCTTGATT	AAATCTTAGA	5880
	GCGAGTTCAG	TTAATCCTTT	AGAATTAACA	TCATAGACAT	TGGTTGCACC	TGGTAAGTGA	5940
	CCAACCTTTG	AAGGATCACC	AGTTATGGCT	AATATTTTCG	TAACGCCAAT	GAGCGATAAT	6000
40	CCAAGTAAAT	GGGACTGCAA	GCCGATTAAG	TTTCGGTCTC	GACATGTAAT	ATGTACGAGT	6060
	GGTTCAATAT	TGTAATATTG	CTTAATTAAG	CTAGCAGCAG	CAATATTGCT	AATTCTGACA	6120
45	GTTGCCAATG	AATTATCTGC	GAGTGTTACC	GCATCTACAT	TAGCTTTATC	AAGTTTAGCG	6180
	ATATTTTCAA	AAAATCTATC	CGTGTCTAAA	TGTTTCGGTG	TATCCAATTC	GATAATAACG	6240
	GTTGGACGTT	CTTGAACCTT	AGATGTTAAT	GATTGTCTAA	CTTTATTTTG	AGATGGATTG	6300
50	AAAAGTGCTT	TCGTTGGTAT	CGGAATCACT	TTTTTGTCAT	TAACAGGTTT	AAGTGTCTGA	6360
	ATAGATTCTT	TAATAAATTT	GATGTGCTCT	GGCGTTGTAC	CACAGCAACC	ACCAATTAAA	6420
55							

	TACTTAAATT CACTATTTTC AATATCTAAT AAGCTGGCAT TTGGATAACA AGATAAGAAT	6540
	GCGTGCTCTG GTAATTC AAT ATGTGTGAAA GACTCTTGCA TATGGTGCGG GCCATGATGA	6600
5	CAATTGAGTC CCACGATGTT TGCACCACAT TGAACGAGTT GTTTTAATCC TTCATTGATT	6660
	GCCTGACCAT TAACTAAGTA ATTTGTGTTT GAAGCGGTGA ATTGAGCAAT GATTGGAATG	6720
	TCGTATTTCT TTCTCGTTCG TGAAATGACA TTTGTAACT CTTCTAGGTC GTAATACGTT	6780
10	TCGAAAAGTA GCGCGTCAAC GCCTTCTTCA ATTAAGGTGT CTATTTGAAT TTCAGTATGA	6840
	TAAAGAATAG TTTGTAAGCT GATATCCTCT TGTTTGATAC CTCTAAACCC ACCAACTGTG	6900
	CCTAATATAT ACGTATCTTT ATTTGCTGCT TTTTTGCGA TGCGAACGGC GGCTTGATGT	6960
15	ATTGCTTTAA CTTTATCTTC AAGACCGAAT CGTTTTAACT TTTCAAAAT TGCACCATAA	7020
	GTATTGGTTT GAATGACATC AGCACC GGCT TCAATATATG AACGATGGAT GCGTTCAACT	7080
	TTATCTGGAT GGCTAAGATT ATATGCTTCT GGACAGGTGT CTAATCCTTC AGAGTATAAA	7140
20	ATGGTTCCTA TAGCGCCATC AGCTACTAAA ACATTATCTT TCAATTGTGT GAGGAATTGA	7200
	CTCATTGAAT GCCTCCTTTA ATGCGTATTT GATGTCTGCA ATGAGTTCAT CAGGATCTTC	7260
25	GAGACCAACA CTTAATCGGA ATAGACCGAA AGTGATACCA CGTTCTTGTC TCACTTCTTC	7320
	AGGTAGTGCA GCGTGAGACA TTGTTGCTGG ATGTGAAAGG ATCGTTTCAA CACCGCCCAG	7380
	ACTCACTGAA ACGAGTGGTA ATGTCAGTGC ATCGACAAAT TGTTGTGCTT TAGACTCATC	7440
30	AGCTAAACGA AAGCCAATAA CGGCACCGCC ATTTT TAGCT TGTTCTAAAT GAGCAGTAGT	7500
	GAGTCCCGGA TAATAAACTT CTGAAATTTT ATCTTGCTTT ATTA AAAATG ACACGATTTT	7560
	TTGAGCGTTT TCGACAGATT GTTTAAATCT GATTGGAAAA GTTTTAAAT GTTTAGCAAG	7620
35	TGTCCAGCTA TCCTGAGCAG ATAACATATT GCCTGTACCA TTTTGTATTA AATAAAGAGC	7680
	GTCACTAATT GCCTCATTAT TAGTTATGAC AGCACCAGCA ATTAAATCGC TATGTCCACT	7740
40	TAAAAATTTT GTAGCACTAT GAATGACAAT ATCAGCGCCA AGTAATAAAG GTGATTGAC	7800
	TAACGGTGTC ATAAATGTAT TGTCCACAGC TACCAGTAGT TCATGCTTTT CGGCTATTTT	7860
	AGAAACAGCT TTGATATCAG TAATTTTAAA ACAGGGATTG GATGGTGTTT CGATATAAAT	7920
45	TAATTTTGTG TTTGATTGAA TGGCACCTTC GATTTGTTTC AGCTTTGTAG TATCTACGGT	7980
	TGTAAATTCA ATATTAAATC GATTC AAAAT TTGCTCAGTG AGGCGAAAAG TACCGCCATA	8040
	TACATCATCG GGTAAGATGA CATGATCACC AGATTTGAAA GTCAAAAGTA CTGCTGAAAT	8100
50	AGCAGCAATA CCTGATGCAA AAGCAAAAGC GAATTTTCCC TGTTCTAATC GTGCTAACTT	8160
	CTCTTCTAAA AGTTCACGGT TAGGGTTGCC CTTCGTGCAT AATCATATTT AACATCGCCA	8220

55

	TCCACACCTC TACGCCAATC GAATATCACT TCTGTCTCTT TTGAAAGTGT CATACAATCT	8340
	CTCCAATCTG AGCTTTATCT AATGCTTGGA TGATATCGCG TTCGATGTCT TCATAATTTT	8400
5	CAACACCTAG TGATAAGCGG ATTAAATACT CATCAATGCC ACGTTTATCT TTTTCAGCAT	8460
	CTGGCATATC AACATGTGTT TGGGTGTAAG GGAAGGTCAC TAATGTTTCA GTACCTCCTA	8520
10	AACTTTCTGC AAAAATGCAA ATGTCTAAAT TTTCTAATAA TTTAGCGACG CTATAGGCCT	8580
	TGTTAAGTCT TAAACTAAGC ATGCCAGTTT GCCCGCTATA TAGTACTTCG TCAATTGCTT	8640
	GAAGTGACTG ACATTTTTTA GCAAGTTTTT TAGCGTTTGA TTGCGCACGC TCAATGCGTA	8700
15	AATGCAAAGT TTTAAGTCCA CGTAACAACA AATAACTATC TATTGGTGAA AGTGTGCGC	8760
	CAGTCATGTT GTGAAAATCA AACAACTGTT GCGCGAGTGA TTCATCTTTG ACGGTTACGA	8820
	CACCTGCTAG TACATCGTTA TGTCCGCCAA TATATTTCTG GGCTGAATGT AAGACTATAT	8880
20	CAGCACCTTC TGCTAGTGGT GTTGAAAGAT AAGGTGTTAA AAAAGTATTG TCGATAATTG	8940
	ACAATAAGCC TTTAGCTTTA CAAAGTTGAT AGTATGGCTT TACATCAATA GCAATCATTT	9000
	GTGGGTTAGA TATTGGTTCA ATGAATAATG CAACTGTTTT ATCAGTGATT TCTTTTTCAA	9060
25	CTTGTTTATA ATCTGTAAAA TCAACGTAAT TAAATTTGAT ATCGTATTGT TGCTCGTAAA	9120
	ATTCAAATAA TCTAAATGTG CCACCATATA AATCGAATGA AACTAAAATT TCATCATGAG	9180
	GTTTAAATAG ATTACATATT AATTGAATGG CTGACATTCC ACTTGATGTA GCGAATGATG	9240
30	CAATACCATG CTCAAGTTTG GCAAAACAGG TTTCAAATGT TGAGCGTGTA GGATTTTTAG	9300
	TACGTGTATA ATCAAAACCT GTCGATTGTC CTAGTTTTGG ATGCTTGTAG GCAGTAGATA	9360
35	AATGGATTGG ATTCGCTATA GCACCGGTTG AATCATCGGT TAATGTGATT TGGGCTAACT	9420
	GTGTATCCTT CATATTAAGA CCCTCCTATA AGAAAAATA AAAAAAGCTT CCGTCCTTCG	9480
	TACCCGAATG AATCGGATAA AAAGGACGAA AGCTTATGTT TCGCGGTACC ACCTTTATTT	9540
40	GTTATTCCAT CGCTGAAATA ACCTTATTCA GTACGCATTA AAAGTAAATA TGCTTACTGA	9600
	ACAATTATCA CAATTAAAGT CAGTAAGTAA GGATATAGTA ATGTGCTATC CCATACTTAT	9660
	TAACAAAAAA TCGTGCGTAA AGAATCCAGT ACGCCATTTA ACATCAATGT TAATACTGTA	9720
45	TCGCTATAAC GGGCGAACCC GTAGACACCT CATATTGGCA TCAACACTCC AAGGCCATTT	9780
	TCAAACACGC TTTCAAATC TTCTCTCAGC TACTAAAGAC TCTCTGTATA AGCAGGGTGT	9840
	GTTTACTTCT CCTCTTTATT GTGTTTACGT TTCATTAAAC TGTTATAAGA TATTAATTAG	9900
50	CTTACAGAGT AAAAAAGAT TTGTCAACAA TTATTCAGAA AATTTTGATT TAAAAGTTAA	9960
	TTTGTGTTGT AAATGTAAAT TGGTATCTTG AAGTTGAAAA ATGAATTATT TTTTAAATAA	10020
55		

TCAAATAAAA AGTGATGTGA GTGAATTGTC AAAAAGTGAA GATCAACGTA TTAATAAAAC 10140  
 AAAAGATGAA CAAATTAAGC AAATAGATAT ATCGGATATC AAACCGAATC CGTATCAGCC 10200  
 5 CCGAAAAACT TTCGATGAAA ATCATTTAAA TGATTGGCA GATTCAATTA AGCAATATGG 10260  
 AATTTTGCAA CCAATTGTGC TTAGAAAAAC AGTTCAAGGT TATTACATTG TAGTTGGTGA 10320  
 AAGAAGGTTT AGAGCTTCGA AAATTGCTGG TCTAAAATAC GTATCAGCGA TTATCAAAGA 10380  
 10 TTTAACAGAT GAAGATATGA TGGAAGTGGC GGTCAATCGA AATTTACAAC GAGAAGACTT 10440  
 AAATGCGATT GAAGAAGCTG AAAGTTATCA ACGTTTGATG ACAGATTTGA AAATTACACA 10500  
 ACAAGAAGTA GCGAAACGAT TGAGTAAGTC GCGCCCGTAT ATAGCGAATA TGTTGAGGTT 10560  
 15 ATTACATTTG CCGAAAAAGA TTGCTGACAT GGTAAGAT GGGCGACTGA CAAGTGACA 10620  
 TGGACGAACG TTATTGGCAA TTAAAGATGA ACAACAAATG CTTAGGTTAG CGAAACGGGT 10680  
 TGTAAAGAA AAGTGGAGTG TCAGATATTT AGAAACCAT GTTAATGAAT TAAAAAATGT 10740  
 20 TTCGTCAAAG TCGGAAACAG ACAAAGTAGA TATACTAAG CCTAAATTTA TAAAGCAGCA 10800  
 AGAACGACAG TTGCGAGAAC AGTATGGTAC CAAAGTAGAT ATATCAATAA AAAAATCGGT 10860  
 25 TGGTAAATC TCATTGAGT TTGATTCACA AGAAGATTTT GTGAGAATAA TTGAACAATT 10920  
 AAATCGTAGG TATGGTAAAT AGTTACACAA TTTTATATAA TAACTCTTTG TGCAAGTGTA 10980  
 AATAATTGT AATCAGTGAC ATTTGATTCT AGAT 11014

30 (2) INFORMATION FOR SEQ ID NO: 92:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6022 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

TCCCCTTATG GAATTCACA TTCTAGTTTA CATAATATAT ATTATAGGAA GTTATATGTG 60  
 TGTAACGCAA AAGGTACCCT ACATCATAAT CATTATCTAA TATCGTCACA TAACTTACTT 120  
 45 ATGCTATAAT CATGGTATTA TATTGTTTGG AGTGATTGTA TGAGATTTGT CTTTGATATT 180  
 GATGGTACGC TTTGTTTGA CGGCCGATTA ATTGACCAGA CTATTATTGA TACATTGTTA 240  
 CAATTACAAC ATGATGGTCA TGAAGTTATA TTTGCATCAG CACGTCCGAT TCGTGATTG 300  
 50 TTGCCAGTTT TACCATCAGT ATTTTCATCAG CACACATTAA TTGGCGCAA TGGTGCTATG 360  
 ATTTACAGC AATCAAAGAT TTCTGTTATC AAACCAATTC ATACTGATAC ATATCATCAT 420

55



	GCTGCACAAC TTGACGCTGn AGAACGCGAT TTTTGAGCGT TTAGATCCAC ATAAGCTGGC	540
	CAGTTGTATT GATGTTGCAA ATATCGACAC GCCAATCAAG AkTATTTTAT TAAATATAGA	600
5	CCCGGCACAA ATTACAAC TAATTAGACGA GCTAGATAAA TACCATCAAG AATTGGAAAT	660
	GATTCAACAT TCAAATGAGT ATAACATTGA TATAACAGCG CAAAATATTA ACAAATATAC	720
	TGCATTACAA TATATATTTG ATGCAGATGT TAAATATATA GCATTTGGTA ATGACCACAA	780
10	TGATATTGTC ATGTTACAAC ATGCTAGTAG TGGCTATATT ATAGGACCAT CAGAAGCATA	840
	CACACACGCA ATATTGAAAC TTGATAAAAT CAAACACATC AATAATAATG CACAAGCTAT	900
	TTGCAAAGTC TTAAAATCAT ATAAATAAAA ACACCCCTAT CAAATGATAA TCATTATCAA	960
15	TCGATAGGGG CTATTTTAAT AAAATTCGTC CTCGAACATT TCTTCCTCTT CATCTAATCC	1020
	AAATAATTCT GCCATTTCTC CATGTTCAAT TAACATGTTT AAATATGCAT CGCGGAGTTC	1080
	TTCTTCACTC ATATCATTAA TCATTTCTTT AAGACTATCA ATCCACATAT TTCTGCGTAA	1140
20	TTGATAGTCT TCTTCAACTT CGTTTAAACAT CATTATATGT TTATTTGCTG CTTCTGGACT	1200
	AGCTGTAAAG AGTAATGCAA TCATATGTTT ACATATCACT CGTCTTCCAT CAGCATGAGG	1260
	ACAATTACAT ATGGATTTTC TAGGATGTTC CATATCAATA TAACAACGAT ATACTTTGTT	1320
25	GCCACTGCCC TTTACTTCAG CCTCATGCTG CGTTTCTGAA AATGATTTTA AGTTAATGAC	1380
	GCATTCACCT TGATAATAAT TAAAGCCTCT TTCTATAGAA CGAATACTTG CAATATCAAG	1440
30	TAATCCCAT T AATGaTACTC CTTTTTATTA TTATTTTAA ATAAAGaAA TAAAATAGAT	1500
	AAGTGTCTAG ATTAAAATAC TTGATTTATC TATATTTTAT AACAAGTCTA GAATTATCGC	1560
	ATTCTTAAAT AACTAATATG AAAATGcTTG CACTAATTct TTTGTATAAG GGTGTCTATC	1620
35	AACATTAAAT AATTCctCTA TTGCAAAATC ATCGACTATC ATGCCATCCT TAAGAACGAT	1680
	AATTCTATTA ACTAAGCGTT GTAACACGGA TAAATCATGA GAAATAACGA TAAAATGATT	1740
	TAAGTTCGTA ATCGTTTGGC CTTTAAATAT ATTGATTACA TTTTGTTCAG CTATAACATC	1800
40	TAAATTTGAA GTTATCTCAT CACATATTAA AACGCGAGGC TGTGCTAATA ACGAACGCAT	1860
	GACATTAAAT CTTTGTAATT GTCCGCCACT CACTTCGCTT GGTAATTTAG TCAATAATTG	1920
	CGCGTTTAAAC TCAAAAGTAG ATAAATGTTG TAATAATAAT TGATCCTGAG CAGTATTATC	1980
45	AGTTAGACCT CTGTAATAAT ATAACGCTTC TTTAATGAG GTCTCAATCG TCCAATCAGG	2040
	GTAAAGCTA GTTAAAGGGT GTTGGAATAA CGGTAACACA GCATTGTCAC TTAAGTAAAT	2100
50	CTCTCCTTTA ACAGGTTTAA ACAAGCCAAG AACCAATGAA GCGAGCGTAC TTTTACCACA	2160
	GCCACTTTTCG CCTAAAATAC CAACATTTTC TCCATCAGGT ATAGTAATAT TGATATCTTG	2220

55

EP 0 786 519 A2

	CCCTCTTTAA TTGTGTTCTA TATTTAATTA GACGTTTCAGT ATACGGATGC AAATGCTCAT	2340
	ACTTGAAATG ATTAATATTA CCTCGTTCAA TGATTTGACC TTCTTTTAAA ACATAAATGT	2400
5	ACTGACAATA TTTCAATACA TGACTTAAGT TATGTGTGAT AATAAATAAT GTTTGACCAT	2460
	GTTCTAATAC AATATGCTGT AATAAATCCA TCACTTGATT ACCGTTCAAA GCATCCAATG	2520
	ATGCAACTGG TTCGTCTGCA ATGATTAATT TAGGCTCCAA CATGAGAACG CTTGCTATGT	2580
10	ATACGCGTTC AAGTTGGCCC CCAGAAAGTT GGAAACTATA TTTATTTAAT ATATCTTTGC	2640
	TTTGTAATTT AACCCACGAC AAAGCCTTAT CAACTTTGGA CAAAGCCTCT TCTTTACTAC	2700
	CTTTATAATG CTTACGATAA ATCGCAGTTA ACTGTTTACC TAATTTAGTA TGGTCGTTAA	2760
15	AACTTTCTGC ATAATTTTGA GAAATATAGC CAATGTATG ACCATAATAT TGACTIONC	2820
	TACTAACATT TTCCCATCA AATTGGTACG AATCATACGT GCAGCTTAAA TCAAATGGTA	2880
	AATATTCAAG TAAAGCTTTA GCAATCAAAC TTTTTCAGC GCCGCTCTCT CCAATCAAGG	2940
20	CATTAATCTG TTGACTAAAA ATTTTCAAAT CAATCCCTTT AATAAGAGAT TTCTCACTAG	3000
	TATTCCTTTAT TGTAAATTT TGTATATCAA TGAGACTCAT CATATTCACC CCGTTGTTTC	3060
25	AGCAATCTAT CTCTTAGTGC ATCACCAGTT AAATTAAAAA TTAAATAGT TATAGCAATG	3120
	ACTGAAGCAG GTGCAATCAA CATAATTGGA TGAGACGAAA TAAATCACG ACCTTGTTGC	3180
	AACATAGCGC CCCaCTCTGG TGTGGCGGT TGTGCACCTA ACCCAATAAA TGATAGTGAA	3240
30	CTTATATATA GAATGATTTT ACCGAAATCA ACGACCATCA AAACGATAAT AGCCGGTATA	3300
	ATTTTAGGTG TTAAATGACG TATTAATATT GTTCTTGTG GTACATGAAA TAATTGTGCC	3360
	ATTTTATAT AAGGCTTATT CATTTGCTA TTAACATAC TTCTAGTCAA CCTTGTGTAA	3420
35	TTCATCCATT TTATTAATGT AATTGAGATA ACTAAATTC ATAAAGATGG TTGAAAAAA	3480
	CTTGCTAAAG CAATCATGAT GATAAATTCT GGAATACTTA GACCAACATC AATAAACCTT	3540
	AACACTAATC GTTCAATCCA CCCTTTTTTG TATCCGGCAA ATAGACCTAG TGTAACACCT	3600
40	ATGACAACGA TAGCTATTAA TGTTAAAACA GTAACAAACA ATGTTGAACG TGCACCGATA	3660
	ATAATTCGGG TAAATAAATC TCTCCATAA TCATCAGTTC CTAATAAATG CAACCAACTA	3720
	ATAGGTTCAA AAGTTTGTGA TAAATTGACT TTGGTTGCAT TTCACTACT GACAAAGAAT	3780
45	TGCAGTACAA TTACCACAAA AATAAATGCA ACGAATACAA AAAATATCAG GTTATTCTTT	3840
	GAAAATATTT TATGCATGAC GGTCACTACT TTCTGATATC AATGGTGTAT TGGTTTTGAT	3900
	TTTTGGATTT CCTAATTGTA AACGCTGCTT CGGATCAAGT AATAACGTTA ATAAATCAGC	3960
50	AATCGTATTG ATAATAACAA CGAAGAAGCC AATAAATAAC ACGCATCCTT GAATAACAGG	4020

55

	ATTTTCAATC ACTACAGTAC CACCTATTAG ACTGCCAAGT GAAATCCCTA GTAATGGGAT	4140
	AATCGGCAAA ATTGTTGGTT TTAGTAAATC ATGAATTAAA ATATAACGTT CATTCATACC	4200
5	GCGTAATCTT GATGCTTGTA CGATATTACT TTGCAATAAC ATCAATAAAT TAGAACGCAC	4260
	TAAACGAATG ATGTATGCAC ACATACCTAA AGATAGCGTG ATTACAGGTA ATATAAACTG	4320
	ACTTAGTATA ACGCTATCTA TATTCATTAA ATTTGTGACA ATAAATAATA AAATAATACC	4380
10	GATAAAGAAC GCTGGTAAAC TAATCGATAG TGTTGAGATC ACTCTAATCA CTTTATCCGT	4440
	CCACTTATGA AATCGTTTGG CTGCTATAAT GCCGAGCGGT ATAGATATGC ATAACGACAC	4500
	TACTAATGTT GAAAATGATA TGAGTAATGT TATGGGTGCA TAGTTGAATA ATATCTGTGT	4560
15	TACCGGTTCT TTTGATTCAA AACTTTTTCC TAAATTAAAA TGTAATAAAT GATTCATCCA	4620
	ATGCCACCAC TGTACCAATA AAGAATCATT TAATCCCAAT TTATCTTTGG TTGCATTTAT	4680
	TTGTTCCGTC GACACTTGTG CTACATCAAG ATGTAATATT TTATCAACAG GATTGCCTGG	4740
20	TGATAATTTT ATTAATAATGA ATGTAAGTGT AGAAATAACA AATAAAACAA CTATCATTTG	4800
	CATCAGTCTA TACAACATAG ACTTTATTAT GAACATAATA GTCCCCCTCC TTGTGTAAGT	4860
25	TACTAACACT TTCTTTTTAC ATGAGAATGG CGCATGTATA TGCAACTTAC ATATTAAGAA	4920
	CTAACGTTCA TTATAGTATT ATCCATAAAG AAATTGAAGT ATATTTAATT TTTTAACAAA	4980
	ATCATTATAA AATATAATAT TTTGAATCAA GTCAACCATG TAAATATAA AAAAGTCAAA	5040
30	ACAAAAACAA CTATAGCACT GTATTCCATC TCTTTCGAAA TAATTGTTAC TGCAGTGTA	5100
	CTTAAAAGTC GATGATTTTG TGCATATAGT TGTCGAATAT TATTTTTTAT CTTTACGGCG	5160
	AAGTTCAGCG CCCTCATAGC CGTATTTTTT AATTTGCTTT TCTAATTTAC GCGCTTTTCT	5220
35	TTCTTTACGC CAATTTCTAG TAAAATACCA TAATAGAAAA CTAATTAATA AACTCATAAT	5280
	CGCTAAAAAT GCAGCGTATC CTAATAATGG TTGATATTTT ATATCTTGAA AATTGGAAT	5340
	AAAAAATGCA AGCACACCTA ATATAACAAA TGTAATTACT GCAGATACAA ACCATTTATT	5400
40	TAAACTAAG CAACAGAATA TTGTTAATAA AATCATTATT AATGTTGTGA TCCATAAATA	5460
	ATTAGGCATA TCGAATAATG TCATATTCAT TCTCCTTTTA TTTCATTACT TTCCTGTAT	5520
	ACATTTTATT ATAAATTTTT AAAAAGTTAA ACAATAGCAG TCAGTTTCAA GCAATATTCT	5580
45	ATCTACTAAT AGAAAAATCA TTGTTCCCTG CGACATGGAA ATCGTAACAT TATCGTTTAG	5640
	GAGACAAAAT TATGTATAAT GAATGTATTA TACCAAAGGA GTGATTATAT GTCTCAAGGT	5700
50	TTACCTTTAA GAGAAGATGT TCCTGTTTCA GAAACATGGG ATTTAGTAGA CTTATTTAAA	5760
	GATGATCAAC AATATTATGA AAGTATTGAC GCTCTAGTAC AACAAGCAAA TCAATTTTCA	5820

55

GAAAATATTT TAATTGCCTT AGATCGCTTA AGTAATTATG CAGAACTACG TTTAAGTGTA 5940  
 GATACTAGTA ATATCGAGGC ACAAGTATTG AGCGCTAAAT TATCTACTAC ATACGGTAAA 6000  
 5 ATTGTTAAGC CAATTATCCT TT 6022

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 476 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

CCATCAATAA TGTATACATG ATTGGCATCA TATTCCCCTT TAATTAGAGA GCTACGTACA 60  
 GTTGTGTTTA TTAAAGTAGA ACTAATAAAT AACCATCTCT TATGTGCACA AACACTTCCC 120  
 20 GCAACAATTG ATTCAGTTTT ACCAACCCGT GGCATACCTC TAATGCCAAT CAACTTATGA 180  
 CCTTCTTCTT TGAACAATTC AGCTAAAAAG TCTACTAACA AGCCTAAATC TTCACGCTCA 240  
 25 AATCGAAAGG TTTTCTTATC TTTTGCATCT TGCTCAATAT ATCTTCCATG TCTTACTGCA 300  
 AGACGGTCTC TTAATTCTGG TTTTITAAGC TTTGTTATTT CAATTTTATT TATACCACGA 360  
 GCTATTTGCT CAAAACGTTT AACTTTTTCA AGATTGCTG TTTTAATTAA AAGGCCTCGT 420  
 30 TTACCTTGAT CAACACCATT AATGTAAACA ATACTTATAC CTAACATACC TAATAA 476

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 3633 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

AGAAATACAA CGAAGCATAT AAATATAACC GATCTTTTTT CTAATTGAAT ATTAAGTAAG 60  
 TGTATGTACT TTCTGGAAGT AGCACCTAGT rGGATTGTtC CTCCTACAAC AGGCCAAAAA 120  
 45 TTTTATTTTT TAACTGGCTT AACAGTGTTT AGTTTTTCAT ACTCTTCTCT ACTAATTTTG 180  
 GCGCACCTTT TTGGAATGAA CCAATTAATA AATGGAAAAA AGTATACAAG CCAAGTTCTT 240  
 ATTACATCGA CCATTAAATA CTCATCATCA TACTTAATAA CTCTGTATTT CGGATTTTTTA 300  
 50 TTGATAATTT CGGTTTCACA AAGCAATAAT TATCACTTCC TATTAATAAC AAATTCACAC 360

55

	TTATATGACC TTAAATATAT AACATGAATC TTTTGTCTA TTATTGAAGA CATATTTATA	480
	AAGAAAAATA GCATTGTCAT AATAACCCAA GCAATAAATA CTATAATATT TTGGATAGAT	540
5	AAACTAATCA rrACATCTAA GAACATGATT gATAATCCAC CACAGAAAA ATAAGAAAAT	600
	AGTACAAAGC AAAGATTCTT GAATGATGGA AAAATCATAA TTTTCCATT GCTACTCCGA	660
10	TCATTATAGA TAGATAACTT TACTTTCTGA TTTAAATATA TATAAACAC TAGAATACTT	720
	AATAATAAAA CCGAACAAAT GATAATAACG CAATTTTTTT CTAAATGAGA ATCAGGTATA	780
	TATATTTTAT CTCTAAACAT AGTGCCAAAT AAAAGTATGC TACCTATAGC TGGCCATAAA	840
15	GCTTTaTTTT TAACTGGTTT GACAATATTT AAATTATCAA AATCTTCTCT GCTGATTTGG	900
	ACATATTTTT TTGGTATTAA CCAATTAATA AACGGAAAGA ACAAACATAA CCAGGTGCTT	960
	ACTAAATCAA TCATCAGATA GTCGTTTTTA TATTTAATAA TTCTATATCT GGGATTTTTG	1020
20	TTTACAAC TCACCTCGCA AAGCAATATC TCCACTTCCG TCTCGTTGGT TTTATATCTA	1080
	ATACACTTTC AGATACTTTA TAAGTGTTTT GTATTTTAGT AACATACTAT TTTCCTGTTT	1140
25	ATTACTTAAC TTACGAACTA CAATCTAAGT TTAGTAATTT CTATTGCTTT TTAAGTTTGG	1200
	CATAAACCTT TTTATTACTA ATTGAGCCCA TGCTTATTAG AAAGAAAAAA ATTGTAATAA	1260
	TAATCCACAT AATAAATACC AGTAGATTTT GAGGTTTTAT AGTCATTAGC CATATTAAAA	1320
30	ATAATATAGA ACAACCTCCT AATAATAGAT ATGTGAAAAC TATAAACTT CCATCTTTAA	1380
	AAGTAGGCAC TAATATAACC CTATTTTCAT TATCTAGATT ATCATCATAT ATCTTTAGTT	1440
	TAAGCTTTTT ATTTAAGTAA ATGTAAATG CTGCAATACC TATAAATCCT ATAAAACATA	1500
35	AAGATATTAA AATCTTATTA TCTAATTGAA CTTCAAACGT ATGTACATAT TTCCGTAAAA	1560
	TAACTACAAA TAAAAACGAA CTACCAGTAA CTGGCCAGAA AATATTATTT TTATTTTGTT	1620
40	TATCAACATT TAAATTTTCA AGTTCCTTCT CACTAAGTTT TGCATACCTT TTGGGAATGA	1680
	ACCAATTAAT AAAAGGAAAA AAGTATACAA GCCAAGTGCT TACTAAATCA ATTAACAAAT	1740
	ACTCATCATT ATATTGAACG ACTTTATATC TCGGATTTTT ATTAATAACC TTAATATTAA	1800
45	AAAGCAAAAC TCACCACGCC CATTTTCATTG GATTTATATG ATTGCTAATA ATATTTTATG	1860
	CTTCACTAAC AGCATTCCCA ACACTATCCA TGGATTTTTT TGTAGTTTTT TTAACAACAT	1920
	CTATACTATT ATCGATTTTA TGCCCTACCC AGTCTACTTT ATCTTTTAAT CCAAAAATAT	1980
50	TATTTTGATA AATTAAATCT GTTCCTAATG CAAATACTGT ACTCATAGCC AAACCTGCTA	2040
	AAATCACCCA TCCTACTGGA TTAATTCCTA AAACAAAAGT CGCTAATCCA GCTCCAAC TG	2100
55	CTGTCCCTGC AGATCCAGCT GCAAGCGTgC ATACCATTAT GCGACAACGC CTCTCCAAAT	2160

CCTTTACCTA GGTATTTTCC GCCTTTTGCA AATTTACTAC CATTTTCTAT AAACACATTA 2280  
 CCTGATGTAC GTTTGACTTC CACAAATGAA TTTGGACCTG CTGGGCCTTT CACTCCACCT 2340  
 5 GCTGTATTGa TAAATACACC GAATTTACTT GcATTTATAC CGTCTTGCTC TAAAAGTGTT 2400  
 GACGTAATAT CTAATCCTAT ATCTCTTTTA ATACTGTCTT TATTGTCATT TATATATTTT 2460  
 AATATACTTT TCGGGATATC GTCTTCTGGA TGTTCCTTGG CATATGCCTT TATAACAGCA 2520  
 10 AAGTCTGCTT TATTTAAAGT TTCTTTCTCT GCTTTATGTT CAATTTTCCC CATAGCAACT 2580  
 TTCAAATATT TTTCACTACT TGCTTTGGCC CAATCAAGTT CTTTACCTGA AGGAATATTA 2640  
 15 AATTGATTGG TTGAAAAGTT CCAAAAATTC TCGCCTTGGG TAAGTCCTTG TTGGACAATT 2700  
 TTTTGAAATT CTTCAACTTC TTTAAATATT TCTGGTGATT TTTGATTAAA CTCACGCAAT 2760  
 TTGCGTAGCT TCTCTTCTAA TTCACTGTTTT TGTGACCTA ATGTTTCGTAT TATTGTGTTG 2820  
 20 TTCGATGAAA TGGCTTGCTG ATTATCGGAA GCATGCTTTT TCAAATTGTT ATTCAAATTT 2880  
 TCATATCGCG TAATTTGTTG ACTTAATGAT CTGATATCTT CTCAAGCTC TGATTCTTTT 2940  
 AAAGATATGC TATCAACCTC ACTCGTATAA CGTGACACAA AATTaTCGCA AGCTTGCTTC 3000  
 25 GTTAAATCAC TCAATGTTTT CATACTTGTT GATAATGGAA TTAACACCGT ACTAAAAAAT 3060  
 TGCTTAGCTG ACGTATACGC TTTCCTTTTA AGCGCATCAT CATTAAATAA TTGAGTAATT 3120  
 GCTTTTTCCTA ACGCATCATA ATTTGAATTC ATTGTTTGAC TCAAATTCCC CACACTTGAA 3180  
 30 GCTTGGTTTC GAGATCTGTC TAAATACATG TCAATACTCA TCGGCATGCT CCTTTTTCAA 3240  
 AAATATATGA TTTTCAAACCT ATTTAAATC AAATGCTTTT TACATCTACA AAGTTGTAAA 3300  
 35 ATTTTAAAAC TCGGCGATGA TTATTCTTTA TGTAAAGGAG TCTAGATGCA GGTAATTTGA 3360  
 GATAACATGT CGCCTTTTIT CTTATTTTAG CATATGGATA TAATGGTGTC TTTGTATATT 3420  
 CGCAATTAAT CAATAAAAAT TATCTTTCAA TATTTTAATT TTATTGCGAC AACATCCTTA 3480  
 40 ACATTAAATA TATTAATATC TCAAAATATA TTCACTATTA AAATATGTCA TCAGTTGTTA 3540  
 AAAGTATTTT CTCATCATGC GAAATATCAA AACGTATCTA AAATACGAAT AAGTTTATAC 3600  
 AATCACACAA CATCATCATT CAAAATTTTA TTG 3633

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2365 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

	TGATACGAAt	GCATTACAAT	TCATATGCAA	CATACAATTc	CTTCTACAGC	AAATGAAGTG	60
	AAACAAATAG	TTGATGTGAC	ATCTGTAGCA	GAAAATGATA	CGCATTAGTC	ATAAAATTAA	120
5	ATGGAAATGT	CGATGAAGTG	TATCAGCAAT	TACAGCGATT	AATTAAGAAT	GCTAATGTCG	180
	AAGAGAGTGA	GAATACTGAC	AATATTAATA	GTCAAGATAC	AAGTTATACA	CCTCAAGTAA	240
	AAGTAACAAC	ACCAATTTTA	GTGAAAGCAC	CAATCGCTGG	TCGTCGTATT	TTACTTAAAG	300
10	AAGTAAGAGA	TTCAATTTTT	AGAGAGAAAA	TGGTAGGTGA	AGGCTTAGCA	ATCAAAGCTC	360
	ATGAAGAATC	CAAAGTAATC	GCACCGTTCA	ATGGTTTAAT	ATCTATGATT	GTACCAACTA	420
15	AGCATGCAGT	TGGTATTCAA	TCAGAAGACG	GTGTGGACAT	AGTCATTcAT	ATTGGCGTGA	480
	ATACAGTTGA	CTTGGAAGGT	AAAGGGTTCA	AGTGCTTTGT	AAAGCAAAAT	GATCATGTTG	540
	AAGCAGGGCA	AACGTTGTTG	CAATTCGACC	AGCAATATAT	ACAACAACAA	GGCTACAATG	600
20	CTGACGTTAT	TGTCGTTATT	AGCAACTCTG	CCGATTTAGG	AAAAGTAGAA	CTGACAATGA	660
	ATGAAATCAT	TACGACTGAA	GATGTTATTT	TTAAAATATT	TAAAAACTAG	GAGTGTGTTG	720
	TAATAATGAC	AAAATTACCG	CAAATTTTCA	TGTGGGGTGG	CGCTCTTGCC	GCAAATCAAT	780
25	TTGAAGGTGG	ATATGATAAA	GGTGGTAAAG	GGTTAAGTGT	AATTGATGTT	ATGACGAGTG	840
	GTGCACATGG	CAAAGCACGT	CAGATTACAG	AATCTATAGA	TCCCAATCAC	TATTATCCAA	900
	ATCATGAAGG	TATTGATTTT	TATCATCGTT	ATAAGGAAGA	TATTGCCTTG	TTTAAAGAAA	960
30	TGGGATTGAA	ATGTTTACGT	ACGTCGATTG	CGTGGACACG	TATCTTTCCG	AATGGGGATG	1020
	AAGATGTGCC	AAATGAAGAA	GGACTCGCCT	TTTATGATCG	TATCTTTGAT	GAATTAATTG	1080
35	CACAAGGTAT	TGAACCTGTT	GTGACGTTAT	CACATTTTGA	GATGCCACTT	CATTTAGCGA	1140
	AACATTATGG	TGGATTTAGA	AATAGAGAAG	TTGTcGATTA	TTTTGTGCAT	TTTGCGCGTG	1200
	TTGTATTTTGA	AAGATATAAA	GATAAAGTTA	CATATTGGAT	GACGTTTAAT	GAAATTAATA	1260
40	ATCAGATGGA	CACATCAAAT	CCTATCTTTT	TATGGACGAA	TTCTGGGGTA	GCATTGACAG	1320
	AAAATGATAA	TCCTGAAGAA	GTcYTGATc	AAGTAGCACA	TCATGAACCT	TTAGCCAGTG	1380
	CyTTAGCAGT	TCGTCTTGGT	AAAGaGATtA	ATCCgAaGTT	TAAGATTGGr	ACmATGATTt	1440
45	CAmaTGTACC	CmTTTTATCCa	TAwTCGTGTC	ATCCGAAAGA	TATGATGGAA	GCACAAATTG	1500
	CGAATCGCTT	ACGTTTCTTT	TTCCCGGATG	TCCAAGTGAG	AGGTTATTAT	CCAAGCTATG	1560
	CTAAAAAAT	GTTGGCACGA	AAAGGATATG	ATGTTGGATG	GCAAGAAGGG	GACGACAGTA	1620
50	TTTTACAGCA	GGGCACGGTT	GATTATATTG	GCTTTAGTTA	TTACATGTCT	ACGGCTGTAA	1680
	AACATGATGT	TGATACTACA	GTTGAAAACA	ACATCGTCAA	CGGTGGTTTG	AATCATTCTG	1740
55							

GATATACATT GAATGTGTTA TATGATCGTT ATCAGTTACC ACTTTTATT GTGGAAAATG 1860  
 GTTTTGGTGC AGTTGATGAA GTGGTAGATG GACATATTCa TGATGATTAT CGCATTGAAT 1920  
 5 ATTTAAAAGC ACATATTACA GCAGCGATAG AAGCAGTTGA TCAAGATGGT GTAGATTTAA 1980  
 TCGGTTATAC ACCGTGGGGA ATCATTGATA TTGTTTCATT TACAACCGGT GAAATGAAGA 2040  
 AACGCTATGG TTTAATATAT GTTGATCGAG ATAATGATGG TCATGGCACG ATGGAACGCT 2100  
 10 TGAAAAAGA TTCGTTCTAT TGGTATCAAC AAGTGATAGC ATCAAATGGA GATAAATTAT 2160  
 AAAGGTATAT TATAAGTATT TTAGGGTTAG AGCCCGAGAC ATAAATTAAT ATAGTAGGAC 2220  
 CTACAGTGTT ATAATGGCGG gCCCCCAACA CAAAGAATTT CGAAAAGAAA TTCTAcAGGT 2280  
 15 aATGCaAGtT GGCGGGGcCC AACACAGAGA AATTCGAAAA GAAATTCTAc AGGTAATGCA 2340  
 AGTTGGGGAA GGACAGAAAT AAATT 2365

20 (2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11050 base pairs

(B) TYPE: nucleic acid

25 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

30 CTGCGATACG ATTTGTTGAA AGTGGGGAAA ACAAAAAAGT TATCATTACC AATTTAGAGC 60  
 AGGCATACGA AGCTTTGATT GGTAATAAAG GTACACACAT TCACATGTAG CACTTTATCA 120  
 35 CGCGACAAAA CATTAAATAT GTTTCTCCGT TGATTCAAAT GAAAAAGTTG TCTGCTGACA 180  
 CTTTGCAAGG TTTGAAGGAG TTAACTTAT GACAGAAAAC TTTATTTTGG GTAGAAATAA 240  
 TAAATTAGAA CATGAATAA AGGCATTAGC AGATTACATT AATATACCAT ATAGTATATT 300  
 40 ACAACCATAT CAAAGTGAAT GTTTTGTGAG ACATTATACG AAAGGCCAAG TTATTTATTT 360  
 TTCGCCACAA GAAAGTAGCA ATATTTACTT TTTAATTGAA GGTAACATTA TTAGAGAACA 420  
 TTACAATCAA AATGGAGATG TATATCGTTA TTTTAATAAA GAGCAAGTAT TATTTCCAAT 480  
 45 CAGTAACTTA TTTCATCCGA AAGAGGTTAA CGAATTGTGT ACAGCATTAA CCGATTGTAC 540  
 AGTTCTTGGA TTGCCTAGAG AATTGATGGC CTTTTTGTGC AAAGCTAATG ATGATATATT 600  
 TTTGACACTT TTTGCATTAA TAAATGATAA TGAGCAGCAA CACATGAACT ATAACATGGC 660  
 50 ATTAACAAGT AAATTTGCTA AAGATCGAAT TATCAAATTG ATATGCCATC TATGTCAGAC 720  
 AGTAGGATAC GATCAAGATG AATTTTATGA AATCAAACAG TTTTAACTA TTCAACtCAT 780

55



	TGAAAAACTT GTTGTTAAAG ATCATAAAAA TTGGTTAGTA AGCAAACATT TATTCAATGA	900
	TGTATGTGTT TAATATACAA TGTAAATGA ATAAGTTGAA CATGAGGTCT AACGTACATT	960
5	TATACGTTAG GCCTTTTTTG CTAGCATGAT GAATAATTTA AAATGTTAGT TAAATTTGAT	1020
	TGTTGAAATT ACAGTAAAAAT TTAAGGTGAT GAAAAATTTA GAACTTCTAA GTTTTTGAAA	1080
	AGTAAAAAAT TTGTAATAGT GTAAAAATAG TATATTGATT TTTGCTAGTT AACAGAAaAT	1140
10	TTTAAGTTAT ATAAATAGGA AGAAAACAAA TTTTACGTAA TTTTTCGA AAAGCAATTG	1200
	ATATAATTCT TATTTTCATTA TACAATTTAG ACTAATCTAG AAATTGAAAT GGAGTAATAT	1260
	TTTTGAAAAA AAGAATTGAT TATTTGTCGA ATAAGCAGAA TAAGTATTCG ATTAGACGTT	1320
15	TTACAGTAGG TACCACATCA GTAATAGTAG GGGCAACTAT ACTATTTGGG ATAGGCAATC	1380
	ATCAAGCACA AGCTTCAGAA CAATCGAACG ATACAACGCA ATCTTCGAAA AATAATGCAA	1440
20	GTGCAGATTC CGAAAAAAC AATATGATAG AAACACCTCA ATTAAATACA ACGGCTAATG	1500
	ATACATCTGA TATTAGTGCA AACACAAACA GTGCGAATGT AGATAGCACA ACAAACCAA	1560
	TGTCTACACA AACGAGCAAT ACCACTACAA CAGAGCCAGC TTCAACAAAT GAAACACCTC	1620
25	AACCGACGGC AATTAAAAAT CAAGCAACTG CTGCAAAAAT GCAAGATCAA ACTGTTCTCTC	1680
	AAGAAGCAAA TTCTCAAGTA GATAATAAAA CAACGAATGA TGCTAATAGC ATAGCAACAA	1740
	ACAGTGAGCT TAAAAATTCT CAAACATTAG ATTTACCACA ATCATCACCA CAAACGATTT	1800
30	CCAATGCGCA AGGAACTAGT AAACCAAGTG TTAGAACGAG AGCTGTACGT AGTTTAGCTG	1860
	TTGCTGAACC GGTAGTAAAT GCTGCTGATG CTAAAGGTAC AAATGTAAAT GATAAAGTTA	1920
35	CGGCAAGTAA TTTCAAGTTA GAAAAGACTA CATTTGACCC TAATCAAAGT GGTAACACAT	1980
	TTATGGCGGC AAATTTTACA GTGACAGATA AAGTGAAATC AGGGGATTAT TTTACAGCGA	2040
	aGTTACCAGA TAGTTTAACT GGTAATGGAG ACGTGGATTA TTCTAATTCA AATAATACGA	2100
40	TGCCAATTGC AGACATTAAA AGTACGAATG GCGATGTTGT AGCTAAAGCA ACATATGATA	2160
	TCTTGACTAA GACGTATACA TTTGTCTTTA CAGATTATGT AAATAATAAA GAAAATATTA	2220
	ACGGACAATT TTCATTACCT TTATTTACAG ACCGAGCAAA GGCACCTAAA TCAGGAACAT	2280
45	ATGATGCGAA TATTAATATT GCGGATGAAA TGTTTAATAA TAAAATTACT TATAACTATA	2340
	GTTCGCCAAT TGCAGGAATT GATAAACCAA ATGGCGCGAA CATTTCTTCT CAAATTATTG	2400
	GTGTAGATAC AGCTTCAGGT CAAAACACAT ACAAGCAAAC AGTATTTGTT AACCCTAAGC	2460
50	AACGAGTTTT AGGTAATACG TGGGTGTATA TTAAAGGCTA CCAAGATAAA ATCGAAGAAA	2520
	GTAGCGGTAA AGTAAGTGCT ACAGATACAA AACTGAGAAT TTTTGAAGTG AATGATACAT	2580
55		

	ACCAATTTAA AAATAGAATC TATTATGAGC ATCCAAATGT AGCTAGTATT AAATTTGGTG	2700
	ATATTACTAA AACATATGTA GTATTAGTAG AAGGGCATTG CGACAATACA GGTAAAGAACT	2760
5	TAAAACTCA GGTATTCAA GAAAATGTTG ATCCTGTAAC AAATAGAGAC TACAGTATTT	2820
	TCGGTTGGAA TAATGAGAAT GTTGACGTT ATGGTGGTGG AAGTGCTGAT GGTGATTGAG	2880
	CAGTAAATCC GAAAGACCCA ACTCCAGGGC CGCCGGTTGA CCCAGAACCA AGTCCAGACC	2940
10	CAGAACCAGA ACCAACGCCA GATCCAGAAC CAAGTCCAGA CCCAGAACCG GAACCAAGCC	3000
	CAGACCCGGA TCCGGATTCTG GATTGAGACA GTGACTCAGG CTCAGACAGC GACTCAGGTT	3060
15	CAGATAGCGA CTCAGAAATCA GATAGCGATT CGGATTGAGA CAGTGATTCA GATTGAGACA	3120
	GCGACTCAGA ATCAGATAGC GACTCAGAAAT CAGATAGTGA GTCAGATTCA GACAGTGACT	3180
	CGGACTCAGA CAGTGATTCA GACTCAGATA GCGATTGAGA CTCAGATAGC GATTGAGACT	3240
20	CAGACAGCGA TTCAGATTCA GACAGCGACT CAGATTGAGA CAGCGACTCA GACTCAGATA	3300
	GCGACTCAGA CTCAGACAGC GACTCAGATT CAGATAGCGA TTCAGACTCA GACAGCGACT	3360
	CAGACTCAGA CAGCGACTCA GACTCAGATA GCGACTCAGA TTCAGATAGC GATTGAGACT	3420
25	CAGACAGCGA CTCAGATTCA GATAGCGATT CGGACTCAGA CAGCGATTCA GATTGAGACA	3480
	GCGACTCAGA CTCGATAGC GATTGAGATT CAGATAGCGA TTCGATTCA GACAGTGATT	3540
	CAGATTGAGA CAGCGACTCA GACTCGGATA GCGACTCAGA CTCAGACAGC GATTGAGACT	3600
30	CAGATAGCGA CTCAGACTCG GATAGCGACT CGGATTGAGA TAGCGACTCA GACTCAGATA	3660
	GTGACTCCGA TTCAAGAGTT ACACCACCAA ATAATGAACA GAAAGCACCA TCAAATCCTA	3720
35	AAGGTGAAGT AAACCATTTCT AATAAGGTAT CAAAACAACA CAAACTGAT GCTTTACCAG	3780
	AAACAGGAGA TAAGAGCGAA AACACAAATG CAACTTTATT TGGTGCAATG ATGGCATTAT	3840
	TAGGATCATT ACTATTGTTT AGAAAACGCA AGCAAGATCA TAAAGAAAAA GCGTAAATAC	3900
40	TTTTTTAGGC CGAATACATT TGTATTGCGT TTTTTGTTG AAAATGATTT TAAAGTGAAT	3960
	TGATTAAGCG TAAAATGTTG ATAAAGTAGA ATTAGAAAGG GGTGATGACG TATGGCTTAT	4020
	ATTTTCATTAA ACTATCATTC ACCAACCAATT GGTATGCATC AAAATTTGAC AGTCATTTTA	4080
45	CCGGAAGATC AAAGCTTCTT TAATAGCGAT ACAACTGTTA AACCATTAAA AACTTTAATG	4140
	TTGTTACATG GATTATCAAG TGATGAAACG ACATATATGA GATATACAAG CATAGAAAGG	4200
	TATGCGAATG AACACAAATT AGCTGTGATT ATGCCCAATG TGGATCATAG CGCATATGCT	4260
50	AACATGGCAT ATGGTCATAG CTATTATGAT TATATTTTGG AAGTGATGA TTATGTTTAT	4320
	CAATATTTT CACTTTCCAA AAAGCGTGAT GACAATTTTA TAGCAGGTCA CTCTATGGGA	4380
55		

	TTATCTGCTG TGTTTGAAGC GCAAAATTTA ATGGATCTAG AGTGGAATGA TTTTTCAAAA	4500
	GAGGCCATAA TTGGCAATCT TTCAAGTGTT AAAGGAACTG AACATGATCC GTATTACTTG	4560
5	CTAGACAAAG CTGTAGCTGA AGATAAACAA ATTCCAAAAT TGCTCATTAT GTGTGGTAAA	4620
	CAAGACTTTT TATATCAAGA CAACCTTAGAT TTTATCGATT ATTTATCAGC CATAAATGTT	4680
	CCTTATCAAT TTGAAGATGG ACCAGGAGAT CATGATTATG CATATTGGGA TCAAGCGATT	4740
10	AAGCGTGCTA TAACATGGAT GGTGAATGAT TAATTATTTT TTGGAAAATA TGTGGCTGCA	4800
	TTAAATACAC AGAGTGAGAG ATACAAACTA TTTACGCACG ACTAACATTT CTAAGTGTTC	4860
15	AAATTATTTT TGTATTAATA TGATTGGCGC AATTGCTGTA TACACAAAAA TGTTTCTCGT	4920
	GAAACTTAGA TTTAGCTTAT AGTTTTATCA TCATTTGTAT GACTTACATT ATAAATTTTA	4980
	TTATAATGAG GTTAACGCTT TGAAAGGAGT CATCATCATG TCGACCAATA AAAACGATTA	5040
20	TGAGCATATG TTGTTTTATT TTGCATATAA AACCTTTATT ACTACCGCTG ATGAAATTAT	5100
	AGAGAAGTAT GGTATGAGTC GTCAGCATCA TCGTTTTTTG TTTTTTATCA ATAAATTACC	5160
	TGGTATTACT ATTAAATCAT TACTAGAAAT ATTAGAAATT TCTAAmCAAG GATCACATGC	5220
25	AACACTTCAA AAATTAAAAG AGCAAGGTCT CATTATTGAA AAAGTTTtag AGACTGATCG	5280
	ACGTGTCAA AAATTATATT CGACGGATAA AGGCGATCAA CTCATTGCTG AATTGAACAA	5340
	GGCGCAAGAT GAATTATTGC AAAATATATA TCAACAAGTC GGTTCGGATT GGTATGATGT	5400
30	GATGGAAGCA TTGGCTAAAG GgCGACCTGG cTTTGATTTT ATTAAGCATT TGAAAGATGA	5460
	AAAAGAAAGC TAGCATCAGA AATGTTAAAA ATCTTCGCAT TCTTAAATTT AAAAAATATG	5520
35	TCAAAAAGTG TATAATAAAA ACATATAATT TAATTGAACT CAGTTTCAAC ACATCTTAGA	5580
	AAGGAGTTTG AATGATGAAA AAATTAGCAG TTATTTTAAC ATTAGTTGGC GGTTTTACT	5640
	TCGcATTAA AAAATACCAA GAACGTGTTA ACCAAGCACC TAACATTGAG TACTAAATTA	5700
40	AACCATAAAA AATTCCCgAA CACCTTGTTA TAGTGCTCGG GAATTTTTTT ATGCTTTACT	5760
	TGAATATATC AAATATTATT TTTGCGCTTT CTGTATTTTC GATATTACCA CTAAATGATT	5820
	CTGATCTAGG TCCGTAAGCG TA <sub>g</sub> GTATTAA CATCCTCGCC TGTATGTCCA TCGGAAGTCC	5880
45	ACCCTGTATA AGATTTATCA TTTACTGGCT TCTGAATAGC GTGTTGTAGG GCTTTTGTTC	5940
	GCGTTTCTAC TTCTGCGGAT TTTTCGTCTT TTTCTTTTTT AAGTAGTCTT TTTAGCTTTT	6000
	TATTCTCTTT TTTAACCTTT TTCATATCAT CTTGTGAAAA TTCAAATCCA TAACCTTCAT	6060
50	TAATAACTTT TTCAGGTCT TCACCTTTAG CCATTTTTTC TGTCATATAT GATCCAGAGT	6120
	GTTTCATAGA TTTAATCGGT TGAGGATTCC ATTCGTATCC TTTATCTTTA CCAATTGTGA	6180
55		

	ATTGAATGGC GTCATCGAAT GCTTTTTC	6300
	TATCGTTGGA ATGCGCTGAT TTATCTATAG	6360
5	TTTTATTGCG CTCAAGCTTA CTAAGTGCAC	6420
	CTTTAGAAGC ATCTATTGCA AGTGGCATAT	6480
	CTTTATCAGA TTTTGATAAC TCCTTACTGT	6540
10	ATTTTTTATC TAAATTGCCA TTACTTTTAC	6600
	CATCAACTTT ATGCTTTCCG TTGATTTTAT	6660
	TATCATCTCT AGAAGTCACG TGTGCAGCAT	6720
15	CTGTTGAAAC AAGACCAGTC GACTTACCTT	6780
	CTTTCTGCTT GTTACTGTCA ACACCGATGG	6840
20	AGGCTGTTCC GCCAGCTGCA GAATCAGTAA	6900
	GATTTGTGCC TTTTAAATAT GAATCAAAAG	6960
	CATTTTTATA ATAACGATAA GCTGTGTAA	7020
25	AGATAACATT TTTTGGATTT TTAGTATTAC	7080
	TATCGGATTG CGCAATTGCA GGTGTGACAG	7140
	TAGCAACTGC AAATTTTGTG GCTTTTAA	7200
30	TGACTATGTC TTCAGTGTA AAGAGGAATT	7260
	ATGTAAACAG AGTGATTTAG AATAACAAA	7320
35	AGCGTTAGAA TAGAAGCGTG TGAAAATATA	7380
	ATACTAGTAA TGTTATCTCC ATTATTAATC	7440
	GAGCGTAAAC GTACGAAGAA AAAGCAACTC	7500
40	AATACAAATG ACACGGAAAG TTCAAATCAA	7560
	AAAGATAACA AAGGATAATT CAATTGAAGG	7620
	TGAAGATGAT TTTGTTATAG CAGAGAGTTT	7680
45	TGTTATTGTC GTTGAACAAT TTGATGATAT	7740
	GCTTGATTG CTAGATATTA ATTTGCCAAC	7800
50	CCGAAAACA TCTAATGTGC CAATTATATT	7860
	AATTATGGCA ATACAAATGG GGGGAGATGA	7920
55	AACGATTGCC AAAATTCAAG CATTATTGAG	7980

	ACAAAACATA	CAGCTATCTT	TGACTGAATT	ACAAATATTA	AAGTTATTAT	TTCAAAATGA	8100
	AGaTAAATAT	GTAAGTAGrA	CTGCTTTAAT	TGaAAAATGT	TGGGaATCAG	AAAACTTCAT	8160
5	AGATGATAAC	ACATTAGCTG	TTAACATGAC	GCGCCTGCTG	AAAAAATTAA	ATACTATTGG	8220
	CGTTAATGAT	TTTATCATTa	CAAAAGAAAA	TGTCGGATAT	AAAGTATAGG	GTGAATGCAA	8280
	TGACCTTTCT	TAAAAGTATT	ACTCAGGAAA	TAGCAATAGT	CATAGTTATT	TTTGCTTTGT	8340
10	TTGGCTTAAT	GTTTTACCTG	TATCATTTGC	CATTAGAAGC	ATATTTACTA	GCACTTGGCG	8400
	TTATTTTATT	ATTATTACTC	ATATTCATAG	GTATTAAATA	TTTAAGTTTT	GTAAAACTA	8460
15	TAAGCCAACA	ACAACAAATT	GAAAACTTAG	AAAATGCGTT	GTATCAGCTT	AAAAATGAAC	8520
	AAATGAATA	TAAAAATGAT	GTAAGAGAGCT	ACTTTTTAAC	ATGGGTACAT	CAAATGAAAA	8580
	CACCCATTAC	TGCAGCACAA	CTGTTACTTG	AAAGAGATGA	GCCTAATGTT	GTTAATCGTG	8640
20	TTCGTCAAGA	GGTTATTCAA	ATTGaTAACT	ATACAAGTTT	AGCACTTAGT	TATTTAAAGT	8700
	TATTAAATGA	AACTTCTGaT	ATTTCTGTCA	CTAAAAATTC	GATTAATAAT	ATCATTGCGC	8760
	CAATTATTAT	GAAATATTCA	ATACAGTTTA	TTGATCAAAA	AACAAAAATC	CATTATGAAC	8820
25	CTTGTCATCA	CGAAGTATTA	ACTGACGTTA	GATGGACCTC	TTTAATGATA	GAACAATTAA	8880
	TAAATAATGC	ACTTAAGTAT	GCGAGAGGTA	AAGATATATG	GATTGAATTT	GATGAGCAAT	8940
	CCAATCAATT	ACACGTAAAA	GATAATGGTA	TCGGTATTAG	TGAAGCGrAC	TTGCCTAAAA	9000
30	TATTTGATAA	GGGCTATTCA	GGTTATAATG	GCCAGCGCCA	AAGTAACTCA	AGTGGGgATTG	9060
	GTTTATTTAT	CGTAAAACAA	ATTTCAACAC	ACACAAACCA	TCCTGTTTCA	GTCGTATCTA	9120
35	AACAAAATGA	GGGTACAACA	TTTACGATTG	AATTTCCAGA	TGAATAAAAA	CTTCAATAT	9180
	TGTAAGTATA	CTAGTAACAT	TTTTTTACTA	ATTTAAATGT	TATTAGTATT	TTTTTGTTTT	9240
	AATaTAGAAC	TAACAAAGAA	ATGAGGTGCA	TGCCATGTTG	CTAGAAGTGn	AACATGTAAA	9300
40	AAAGGTTTAT	GGTAAAGGTT	TGAATGCTAC	GACAGCACTT	AATCAAATGA	ATTTATCAGT	9360
	TGGAGCTGGT	GaATTtGTTG	CaATTATGGG	TGAGTCTGGG	tCAGGGAAGT	CTACACTACT	9420
	AAATTTAATT	GCTCTTTTTG	ATGGACTAAC	TGAAGGTGAC	ATTATTGTGG	ATGGCGCACA	9480
45	TTTAAATAAT	ATGAAAAATA	AAAGTAAAGC	ATTGTATCGT	CaACAAATGG	TAGGTTTTGT	9540
	TTTTCAAGAT	TTTAATCTTT	TACCAACAAT	GACGAATAAA	GAAAATATAA	TGATGCCATT	9600
	AATTTTAGCT	GGTGCTAAAC	GAAAAGATAT	AGAACAAAGG	GTACATCAGT	TGGCAGTACA	9660
50	ATTACATTTA	GAGGGATTCT	TAAACAAGTA	TCCTTCTGAA	ATCTCTGGGG	GTCAGAAGCA	9720
	ACGCATTGCC	ATTGCACGTG	CATTAGTTAC	TAAGCCGACG	ATTTTACTAG	CCGATGAACC	9780
55							

TCAATTGGAA CAGACAATTT TAATGGTAAC TCATTCAAAT ATCGATGCGT CTTATGCAGA 9900  
 GCGAGTCATT TTTATTAAAG ATGGGCGTCT ATATCATGAA ATATATCGTG GTGAAGAAAG 9960  
 5 TCAATTAGCT TTTCAACAAC GAATAACAGA TAGCTTAGCA CTTGTGAATG GAGGAAGTGT 10020  
 CAATATATGA AGTTAAGATT GTTATGnACA TAGTGCGACG TCAATTTATT ACGCAGCGAC 10080  
 TTGTAATCAT TCCATTCATT TTAGCGGTAA GTGTACTATT CATGATTGAA TATACGCTTG 10140  
 10 TGTCAATTGG GTTAAATAGC TACATAAAAC AGAAGAATGA CTTCTAGTA CCATTTATTA 10200  
 TCATAGCTAA TTTTTTTATG GCGCTTTTAA CTTTTATTTT TATTTTCTAT GCAAATCACT 10260  
 TTATGATGTC ACAAAGACGA AAAGAGTTTA GCATTTTTAT GACATTGGGC ATGACCAAGA 10320  
 15 AAAGTATGCG TTTAATTGTA GTGATGGAAA CTATCTTACA ATTTGTGATA ATTCAGTCG 10380  
 TTAGTATGTC CGGCGGATAC TTACTTGGTG CGATATTTTT CTGTTTATA CAGAAAATAA 10440  
 20 TGGGCAGTGA AGTTGCGACG TTAAGGTATT ATCCATTTGA CTCTGTAGCG ATGTTTATTA 10500  
 CTTTGATTAT CATTGCTGTA TTAATGGGCA TGCTACTTAT ATTCAACTTG TTTAGTATTA 10560  
 ATTTTCAACG GCCGATAACT TATCAACATC GTTCCGATTC TAGTGTCATA TCACGATGGT 10620  
 25 TGCCTTACGT TTTAATTGTT ATAGGAAGCG CAnACTATAT TTAGGTTACT TTATTGCATT 10680  
 ACAACAAGAT ACGACGTTTG GTGCCTTTTT TAAATATGG ATTGTCATAG GATTAGTTAT 10740  
 TATCGGTACT TATGCATTTT TTGTAGGTAT AAGTGAAATA ATTATTAGTA TATTGCAGCA 10800  
 30 GGTATCAAAA GTTTACTATC ATCCACGTA TTTTTTTGTG GTAGTTGGGA TCGTGTACG 10860  
 TCTTAAAATG AATGCAGTCA GTCTTGCAAC AATCACTTTG CTGTGTACAT TTTTGATTGT 10920  
 AACGCTCACA ATGACATTAA CAACCTATCG TGATATGAAT CATACCATTA CGAAATTGAT 10980  
 35 TACGAATGAT TakGATTTGT CATTTAGCGA CAATTCTAAG TCACAAaTAG AACGTCAACA 11040  
 AACaATTGAG 11050

40 (2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 983 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

CGACATAACG AGGCAAGGGT ACATGATACT TTAGCCTCGT TTTTGATATG TATTTTCTG 60  
 AATATAAGGG CAATAGATGG TATTTTATAw TTTTTTTAAG GTAGTGATTA ACATAGATAT 120

55

TCAAGCGGAA CAGCATTATG CACCAGTATT AACGCATTTT TTAGATCCAA GAGGGCAATA 240  
 TATATTGGAA GTGATTTGTG GCAGTTATGA AGATTTAAAC GTATCTTTTT ATGGTGGACC 300  
 5 TAATGCTGAA AGAAAAAGAG CAATCATTTT GCCGAAGTAT TATGAACCTA AAGAAAGCGA 360  
 CTTTGAATTA ACTTTAATGG AAATAGATTA TCCTGAAAAA TTCGTCACTT TAAAACATCA 420  
 ACATATTTTA GGGACATTAA TGTCTTTAGG TATCGAACGC GAACAAGTTG GAGATATAAT 480  
 10 TGTGaaATGAA CGAATTCAAT TTGTTTTGAC AAGTAGATTG GAATCATTTA TTATGTTAGA 540  
 ATTACAACGT ATTAAAGGCG CATCAGTTAA ACTTTATACT ATTCCAGTAA CAGATATGAT 600  
 ACAATCTAAT GAGAATTGGA AAAATGAAAG TGCaCAGTTA GTTCTTTAAG GTTAGATGTT 660  
 15 GTTATTAAAG AAATGATACG TAAATCACGT ACGATTGCGA AACAACTAAT CGAAAAAAAA 720  
 CGTGTTAAAG TGAATCACAC TATTGTTGAT TCAGCAGATT TTCAATTACA AGCAAATGAT 780  
 20 TTAATATCCA TCCAAGGTTT TGGTAGAGCA CACATTACTG ACTTAGGTGG TAAAACTAAA 840  
 AAAGATAAAA CGCACATTAC CTATAGAACA TTATTCAAAT AGTAATGATT TAAGGAGGAT 900  
 AACAAATGCC TTTTACACCA AATGaaATTA AGAATAAAGA GTTTTCACGT GTaaAGAATG 960  
 25 GTTTTAGAAC CTACTGnAGT TGG 983

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:  
 30 (A) LENGTH: 10322 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

TTTTGCAAAG CTTATTTTAT GTCAAACAGA TAGTCAATGT GAAACAAAGG TTAGTACATA 60  
 40 TAATCATCCA GACTTTATGT ATATATCAAC AACTGAGAAT GCAATTAAGA AAGAACAAGT 120  
 TGAACAACCT GTGCGTCATA TGAATCAACT TCCTATAGAA AGCACAAATA AAGTGACAT 180  
 CATTGAAGAC TTTGAAAAGT TAACTGTTCA AGGGGAAAAC AGTATCTTGA AATTTCTTGA 240  
 45 AGAACCACCG GACAATACGA TTGCTATTTT ATTGTCTACA AAACCTGAGC AAATTTTAGA 300  
 CACAATCCAT TCAAGGTGTC AGCATGTATA TTTCAAGCCT ATTGATAAAG AAAAGTTTAT 360  
 AAATAGATTA GTTGAACAAA ACATGTCTAA GCCAGTAGCT GAAATGATTA GTACTTATAC 420  
 50 TACGCAAATA GATAATGCAA TGGCTTTAAA TGAAGAATTT GATTTATTAG CATTAAGGAA 480  
 ATCAGTTATA CGTTGGTGTG AATTGTTGCT TACTAATAAG CCAATGGCAC TTATAGGTAT 540

55

	GAATGGTTTC TTCGAAGATA TCATACATAC AAAGGTAAAT GTAGAGGATA AACAAATATA	660
	TAGTGATTTA AAAAATGATA TTGATCAATA TGCGCAAAAG TTGTCGTTTA ATCAATTAAT	720
5	TTTGATGTTT GATCAACTGA CGGAAGCACA TAAGAAATTG AmTCAAAATG TAAATCCAAC	780
	GCTTGTATTT GAACAAATCG TAATTAAGGG TGTGAGTTAG ATGCCAAATG TAATAGGTGT	840
10	TCAGTTTCAA AAAGCGGGAA AATTAGAATA TTATACACCT AATGATATAC AAGTAGATAT	900
	AGAAGACTGG GTAGTTGTCTG AATCTAAAAG AGGCATAGAG ATAGGTATTG TTAAAAATCC	960
	ATTAATGGAT ATTGCTGAAG AGGATGTTGT GTTACCTCTT AAAAATATTA TTCGCATTGC	1020
15	TGATGACAAA GATATTGATA AATTTAATTG TAATGAACGA GATGCTGAAA ATGCATTAAT	1080
	ACTATGTAAA GACATTGTAA GAGAACAAGG TTTGGACATG CGTTTAGTCA ATTGCGAATA	1140
	TACATTAGAT AAATCGAAAG TTATTTTAA TTTTACGGCG GATGATCGTA TTGATTTTAG	1200
20	AAAATTAGTA AAAATATTAG CGCAACATTT AAAACACGT ATCGAGTTGA GACAAATTGG	1260
	TGTAAGGGAT GAAGCCAAAT TGCTTGGCGG TATCGGACCT TGTGGTAGGT CGTTATGTTG	1320
	TTCTACATTT TTAGGGGATT TTGAACCAGT ATCGATTAAG ATGGCTAAGG ATCAAAATTT	1380
25	ATCATTAAT CCAACTAAAA TTTCTGGTGC ATGTGGTCGT TTGATGTGTT GTTTAAATA	1440
	TGAAAATGAC TATTATGAGG AAGTACGTGC ACAATTACCT GATATTGGTG AAGCAATTGA	1500
30	AACGCCTGAT GGTAACGGGA AAGTAGTTGC TTTAAATATA TTAGACATTT CTATGCAGGT	1560
	GAAGCTTGAG GGACATGAAC AGCCACTTGA ATATAAATTA GAAGAAATAG AAACATATGCA	1620
	TTAAGGAGGC ATTATTACAT TTGGATCGCA ATGAAATATT TGAAAAATA ATGCGTTTAG	1680
35	AAATGAATGT CAATCAACTT TCAAAGGAAA CTTCAGAATT AAAGGCACCT GCAGTTGAAT	1740
	TAGTAGAAGA AAATGTAGCG CTTCAACTTG AAAATGATAA TTTGAAAAAG GTGTTGGGCA	1800
	ATGATGAACC AACTACTATT GATACTGCGA ATTCAAAACC AGCAAAAGCT GTGAAAAGC	1860
40	CATTACCAAG TAAAGATAAT TTGGCTATAT TGTATGGAGA AGGATTTTAT ATTTGTAAAG	1920
	GCGAATTATT TGGAAACAT CGACATGGTG AAGATTGTCT GTTCTGTTTA GAAGTTTTAA	1980
	GTGATTAATC AAGCACACTC AAATAGTGTT ATAATTATAA ATGAATATGG TTTGGATAAG	2040
45	TCTGAGACAA TGCATGTTTC AGGCTTTAAT TGTGTATAAA GTTTTGGTGA TTGCATAAGA	2100
	GATGGCGGTA CTAAATGTTA TTATTAAGTG TGCACGCagT ATCaTTAGTT ATAAAATGTA	2160
50	GCTGTTAAAA GTCAAAAATA CATCGAATGT AGTTAGGCAT ATAATATAAA AAGAGTTTTTC	2220
	AATTACTCAA TAGAAAAAGG TTGTCTTCAT AGGAGTTAAA AATGTTAAAA GAGAATGAAC	2280
	GATTTGATCA ACTAATCAAA GAAGATTTTA GTATTATTCA AAATGATGAT GTTTTTTCAT	2340
55		



	TGGACTTATG TTCAGGCAAT GGGGTGATAC CCTTGTATT GTTTGC AAA CATCCACGAC	2460
	ATATAGAAGG TGTTGAGATT CAAAAACAC TTGTCGATAT GGCGCGACGC ACATTTCAAT	2520
5	TCAATGATGT TGATGAATAT TTAACAATGC ATCACATGGA TTTGAAAAAC GTTACTAAAG	2580
	TATTTAAACC TTCACAATAT ACTTTAGTAA CGTGTAATCC GCCTTATTTT AAAGAGAATC	2640
10	AGCAACACCA ACATCAAAAA GAAGCACATA AGATAGCGAG ACATGAGATT ATGTGTACAC	2700
	TTGAAGATTG CATGATTGCA GCCCGTCATT TATTAAAAGA AGGTGGCAGG CTAAACATGG	2760
	TACATCGTGC AGAGAGACTA ATGGATGTCT TGTTTGAAAT GAGAAAAGTG AATATTGAAC	2820
15	CTAAGAAAGT CGTTTTTATA TATAGTAAAG TAGGGAAATC AGCACAACG ATAGTAGTAG	2880
	AAGGTCGAAA AGGTGGAAAT CAAGGTTTAG AAATCATGCC CCCATTTTAT ATTTATAATG	2940
	AAGATGGTAA TTATAGCGAA GAAATGAAGG AAGTATATTA TGGATAGTCA TTTTGTATAT	3000
20	ATTGTAAAAT GTAGTGATGG AAGTTTATAT ACAGGATACG CTAAAGACGT TAATGCACGT	3060
	GTTGAAAAAC ATAACCGAGG TCAAGGAGCC AAATATACGA AAGTAAGACG TCCGGTGCAT	3120
	TTAGTTTATC AAGAAATGTA TGAGACAAAG TCTGAAGCAT TGAAGCGTGA ATATGAAATT	3180
25	AAACTTATA CCAGACAAAA GAAATTGCGA TTAATTAAGG AGCGATAGTA TGGCTGTATT	3240
	ATATTTAGTG GGCACACCAA TTGGTAATTT AGCAGATATT ACTTATAGAG CAGTTGATGT	3300
30	ATTGAAACGT GTTGATATGA TTGCTTGTGA AGACACTAGA GTAAGTAGTA AACTGTGTAA	3360
	TCATTATGAT ATTCCAACCTC CATTAAAGTC ATATCACGAA CATAACAAGG ATAAGCAGAC	3420
	TGCTTTTATC ATTGAACAGT TAGAATTAGG TCTTGACGTT GCGCTCGTAT CTGATGCTGG	3480
35	ATTGCCCTTA ATTAGTGATC CTGGATACGA ATTAGTAGTG GCAGCCaGAG AAGCTAATAT	3540
	TAAAGTAGAG ACTGTGCCTG GACCTAATGC TGGGCTGACG GCTTTGATGG CTAGTGGATT	3600
	ACCTTCATAT GTATATACAT TTTTAGGATT TTTGCCACGA AAAGAGAAAG AAAAAAGTGC	3660
40	TGTATTAGAG CAACGTATGC ATGAAAATAG CACATTAATT ATATACGAAT CACCGCATCG	3720
	TGTGACAGAT ACATTAAAA CAATTGCAAA GATAGATGCA ACACGACAAG TATCACTAGG	3780
	GCGTGAATTA ACTAAGAAGT TCGAACAAAT TGTAAGTATGAT GATGTAACAC AATTACAAGC	3840
45	ATTGATTCAG CAAGGCGATG TACCATTGAA AGGCGAATTC GTTATCTTAA TTGAAGGTGC	3900
	TAAAGCGAAC AATGAGATAT CGTGGTTTGA TGATTATCT ATCAATGAGC ATGTTGATCA	3960
50	TTATATTCAA ACTTCACAGA TGAAACCAAA ACAAGCTATT AAAAAAGTTG CTGAAGAACG	4020
	ACAACTTAAA ACGAATGAAG TATATAATAT TTATCATCAA ATAAGTTAAT CACTTTATCG	4080
	ATTaTATGAA ATTTTAAACG ATTTTATAAA CGCAAGCTGT AATTTTAAAT GGTAAGTTAT	4140
55		

	GT TTTT TTAAT GT AAAA TAAA TACATTGAAA GTAATAAATA CCTTAACATT GAATAAGATG	4260
	AAAA TGAGAT GACGAGATAA ATGTT CGCGT CCGTTGAAAT GCATAGAAAT CTTAGATATT	4320
5	ATT TGAAGTG AGACATTACG AGGAGGAACA GTTATGGCTA AAGAAACATT TTATATAACA	4380
	ACCCCAATAT ACTATCCTAG TGGGAATTTA CATATAGGAC ATGCATATTC TACAGTGGCT	4440
10	GGAGATGTTA TTGCAAGATA TAAGAGAATG CAAGGATATG ATGTT CGCTA TTTGACTGGA	4500
	ACGGATGAAC ACGGTCAAAA AATTCAAGAA AAAGCTCAAA AAGCTGGTAA GACAGAAATT	4560
	GAATATTTGG ATGAGATGAT TGCTGGAATT AAACAATTGT GGGCTAAGCT TGAAATTTCA	4620
15	AATGATGATT TTATCAGAAC AACTGAAGAA CGTCATAAAC ATGTCGTTGA GCAAGTGTTT	4680
	GAACGTTTAT TAAAGCAAGG TGATATCTAT TTAGGTGAAT ATGAAGGTTG GTATTCTGTT	4740
	CCGGATGAAA CATACTATAC AGAGTCACAA TTAGTAGACC CACAATACGA AAACGGTAAA	4800
20	ATTATTGGTG GCAAAAGTCC AGATTCTGGA CACGAAGTTG AACTAGTTAA AGAAGAAAGT	4860
	TATTTCTTTA ATATTAGTAA ATATACAGAC CGTTTATTAG AGTTCTATGA CCAAAATCCA	4920
	GATTTTATAC AACCACCATC AAGAAAAAAT GAAATGATTA ACAACTTCAT TAAACCAGGA	4980
25	CTTGCTGATT TAGCTGTTTC TCGTACATCA TTTAACTGGG GTGTCCATGT TCCGTCTAAT	5040
	CCAAAACATG TTGTTTATGT TTGGATTGAT GCGTTAGTTA ACTATATTTT AGCATTAGGC	5100
30	TATTTATCAG ATGATGAGTC ACTATTTAAC AAATACTGGC CAGCAGATAT TCATTTAATG	5160
	GCTAAGGAAA TTGTGCGATT CCACTCAATT ATTTGGCCTA TTTTATTGAT GGCATTAGAC	5220
	TTACCGTTAC CTAAAAAGT CTTTGCACAT GGTGGAATTT TGATGAAAGA TGGAAAAATG	5280
35	AGTAAATCTA AAGGTAATGT CGTAGACCCT AATATTTTAA TTGATCGCTA TGGTTTAGAT	5340
	GCTACACGTT ATTATCTAAT GCGTGAATTA CCATTTGGTT CAGATGGCGT ATTTACACCT	5400
	GAAGCATTTG TTGAGCGTAC AAATTTGAT CTAGCAAATG ACTTAGGTAA CTTAGTAAAC	5460
40	CGTACGATTT CTATGGTTAA TAAGTACTTT GATGGCGAAT TACCAGCGTA TCAAGGTCCA	5520
	CTTCATGAAT TAGATGAAGA AATGGAAGCT ATGGCTTTAG AAACAGTGAA AAGCTACACT	5580
	GAAAGCATGG AAAGTTTGCA ATTTTCTGTG GCATTATCTA CGGTATGGAA GTTTATTAGT	5640
45	AGAACGAATA AGTATATTGA CGAAACAACG CCTTGGGTAT TAGCTAAGGA CGATAGCCAA	5700
	AAAGATATGT TAGGCAATGT AATGGCTCAC TTAGTTGAAA ATATTCGTTA TGCAGCTGTA	5760
50	TTATTACGTC CATTCTTAAC ACATGCGCCG AAAGAGATTT TTGAACAATT GAACATTAAC	5820
	AATCCTCAAT TTATGGAATT TAGTAGTTTA GAGCAATATG GTGTGCTTAA TGAGTCAATT	5880
	ATGGTTACTG GGCAACCTAA ACCTATTTTC CCAAGATTGG ATAGCGACGG AaAATTGCAT	5940
55		

	AACCTCAAAT	TGATATTAAA	GACTTTGATA	AAGTTGAAAT	TAAGGCAGCA	ACGATTATTG	6060
	ATGCTGAACA	TGTTAAGAAG	TCAGATAAGC	TTTTAAAAAT	TCAAGTAGAC	TTAGATTCTG	6120
5	AACAAAGACA	AATTGTATCA	GGAATTGCCA	AATTCTATAC	ACCAGATGAT	ATTATTGGTA	6180
	AAAAAGTAGC	AGTTGTTACT	AACCTGAAAC	CAGCTAAATT	AATGGGACAA	AAATCTGAAG	6240
10	GTATGATATT	ATCTGCTGAA	AAAGATGGTG	TATTAACCTT	AGTAAGTTTA	CCAAGTGCAA	6300
	TTCCAAATGG	TGCAGTGATT	AAATAACTGT	ATTTTAAAAA	ATTAGGAGAG	ATAATTATGT	6360
	TAATCGATAC	ACATGTCCAT	TTAAATGATG	AGCAATACGA	TGATGATTTG	AGTGAAGTGA	6420
15	TTACACGTGC	TAGAGAAGCA	GGTGTGTGATC	GTATGTTTGT	AGTTGGTTTT	AACAAATCGA	6480
	CAATTGAACG	CGCGATGAAA	TTAATCGATG	AGTATGATTT	TTTATATGGC	ATTATCGGTT	6540
	GGCATCCAGT	TGACGCAATT	GATTTTACAG	AAGAACACTT	GGAATGGATT	GAATCTTTAG	6600
20	CTCAGCATCC	AAAAGTGATT	GGTATTGGTG	AAATGGGATT	AGATTATCAC	TGGGATAAAT	6660
	CTCCTGCAGA	TGTTCAAAAG	GAAGTTTTTA	GAAAGCAAAT	TGCTTTAGCT	AAGCGTTTGA	6720
	AGTTACCAAT	TATCATTTCAT	AACCGTGAAG	CAACTCAAGA	CTGTATCGAT	ATCTTATTGG	6780
25	AGGAGCATGC	TGAAGAGGTA	GGCGGGATTA	TGCATAGCTT	TAGTGGTTCT	CCAGAAATTG	6840
	CAGATATTGT	AACTAATAAG	CTGAATTTTT	ATATTTTCATT	AGGTGGACCT	GTGACATTTA	6900
30	AAAATGCTAA	ACAGCCTAAA	GAAGTTGCTA	AGCATGTGTC	AATGGAGCGT	TTGCTAGTTG	6960
	AAACCGATGC	ACCGTATCTT	TCGCCACATC	CGTATAGAGG	GAAGCGAAAT	GAACCGGCGA	7020
	GAGTAACTTT	AGTAGCTGAA	CAAATTGCTG	AATTAAAAGG	CTTATCTTAT	GAAGAAGTGT	7080
35	GCGAACAAAC	AACTAAAAAT	GCAGAGAAAT	TGTTTAATTT	AAATTCATAA	AGTTAAAAGT	7140
	GAGAAAGATC	ACCGCCATAA	ATGTAAACGA	TGCTATATTC	GTTTAATATG	CTATGGTTCT	7200
	TTCTCACTTT	TTTAAATTAA	AATATCGTGC	ATGTGGAATA	CGTGCGATAG	AGATGGTTAG	7260
40	AGCTTTGAAA	TTAAGAATTG	TAGGAAGGCG	TTTTAAATGA	AAATCAATGA	GTTTATAGTT	7320
	GTAGAAGGAC	GAGATGATAC	TGAGCGTGTT	AAACGAGCTG	TTGAATGTGA	TACGATTGAA	7380
	ACGAATGGTA	GTGCCATCAA	CGAACAAACT	TTAGAAGTAA	TTAGAAATGC	TCAACAAAGT	7440
45	CGAGGCGTTA	TTGTATTAAAC	AGATCCAGAT	TTCCCAGGAG	ATAAAATTAG	AAGTACAATT	7500
	ACTGAACATG	TCAAAGGTGT	TAAACATGCG	TATATTGATA	GAGAAAAAGC	TAAAAATAAA	7560
50	AAAGGGAAAA	TTGGTGTTGA	ACATGCCGAC	TTAATTGATA	TTAAAGAAGC	GTTAATGCAT	7620
	GTTAGTTCAC	CCTTTGATGA	AGCTTATGAA	TCAATTGATA	AATCTGTGCT	AATAGAGTTG	7680
	GGGTTAATTG	TTGGGAAAGA	TGCAAGGCGC	CGTAGAGAAA	TTTTAAGTAG	AAAATTGCGA	7740
55							

	GCGGATGTAA GGCAAGCTTT AGAAGATGAA TGAGGAAGTG AAAATGTTGG ATAATAAAGA	7860
	TATTGCAACA CCATCAAGAA CGCGAGCGTT GTTAGATAAA TATGGCTTTA ATTTTAAAAA	7920
5	AAGTTTAGGA CAGAACTTTT TGATAGATGT GAATATCATT AATAATATCA TTGATGCAAG	7980
	TGATATTGAT GCACAACTG GGGTGATTGA AATTGGTCCA GGCATGGGGT CATTGACAGA	8040
10	ACAATTGGCC AGACATGCTA AAAGAGTATT GGCATTGAA ATTGATCAAC GTTTAATACC	8100
	TGTATTAAAT GATACACTAT CACCTTATGA TAATGTGACG GTGATTAATG AAGATATTTT	8160
	AAAAGCGAAT ATTAAAGAAG CTGTTGAAAA TCATTTACAA GATTGTGAAA AAATAATGGT	8220
15	TGTTGCAAAC CTGCCGTACT ATATTACGAC GCCAATTTTA TTAAATTTGA TGCAACAAGA	8280
	TATACCAATT GATGGCTACG TGGTGATGAT GCAAAAAGAA GTGGGCGAAC GCTTAAATGC	8340
	TGAAGTAGGT TCAAAAGCAT ATGGTTCGTT ATCAATTGTC GTACAATACT ATACAGAGAC	8400
20	TAGTAAAGTA TTAACGGTAC CTAAATCTGT ATTTATGCCA CCACCTAATG TTGATTCAAT	8460
	AGTTGTAAAA CTGATGCAGA GAACTGAACC GTTAGTAACA GTAGATAACG AGGAAGCATT	8520
	CTTTAAGTTA GCAAAAGCAG CATTTGCACA AAGAAGAAAG ACAATTAACA ATAACATCA	8580
25	AAATTATTTT AAAGATGGTA AACAACACAA AGAAGTGATT TTACAATGGT TGGAACAAGC	8640
	AGGTATTGAT CCAAGACGTC GCGGTGAAAC GCTATCTATT CAAGATTTTG CTAAATTGTA	8700
30	TGAAGAAAAG AAAAAATTCC CTCAATTAGA AAATTAAATG ATTGACAAAG CAAAGCACTA	8760
	TTGTTAAAAT TTAAATTTTG TTTGACGAAA ACGTTGCAAA TATGGTATTA TGTAAGTTGT	8820
	AGCGAGGTGG AGCAATATGC CAAAATCAAT TTTGGACATC AAAAAATTCTA TTGATTGTCA	8880
35	TGTAGGAAAT CGTATTGTAC TGAAAGCCAA TGGAGGCCGT AAGAAACAA TAAACGTTT	8940
	TGGAATTTTA AAAGAAACAT ATCCGTCAGT TTTTATTGTT GAGTTAGATC AAGACAAACA	9000
	CAACTTTGAG AGAGTATCTT ATACATACAC TGATGTGTTA ACTGAAAATG TTCAAGTTTC	9060
40	ATTTGAAGAG GATAATCATC ACGAATCAAT TGCACACTAA ATAAGACATA TAGAGATGTT	9120
	AGACGTTTCT TAGTATAAGA AGTAAATATT ATGATAATTA TTTGAGTGTT GGGCATTATG	9180
	TTCAATACTC TTTTATTTTA CAAAATGTTT AACACTGATG TTTGCTTAT AGATTTTTC	9240
45	GTAAATGGAT AATTGTATTT ATAAACACAA ATACAAGTAA AACTAAGTA ATTAGATGGA	9300
	GAAAATTACT TTTTATTAA AAAAACTA AAAAACAAAT TAAATGTCA AATATTAATT	9360
50	CTCTTTATGT TAAATCATC ATATTAAGAT AACGAAAAGA GGGCGGAAAA TGATATATGA	9420
	AACGGCACCA GCCAAAATTA ATTTTACGCT CGATACACTT TTTAAAAGAA ATGATGGCTA	9480
	TCATGAGATT GAAATGATAA TGACAACAGT TGATTTAAAT GATCGTTTAA CTTTTCATAA	9540
55		

AAATCTCGCA TATCGTGCAG CGCAACTATT TATTGAGCAA TATCAACTAA AGCAAGGTGT 9660  
 AACAAATTTCT ATCGATAAAG AAATACCTGT TTCTGCTGGC TTAGCTGGAG GTTCGGCTGA 9720  
 5 TGCAGCAGCA ACGTTAAGAG GATTGAATCG ACTTTTGTAT ATAGGGGCGA GTTTGGAAGA 9780  
 ATTGGCTCTA CTAGGCAGTA AAATCGGGAC AGATATTCCG TTTGTATTT ATAATAAAAC 9840  
 TGCATATGT ACTGGAAGAG GAGAGAAAAT CGAGTTTTTA AATAAACAC CTTCAGCTTG 9900  
 10 GGTGATTCTT GCTAAACCAA ACTTAGGCAT ATCATCACCA GATATATTTA AGTTGATTAA 9960  
 TTTAGATAAG CGTTACGACG TACATACGAA AATGTGTTAT GAGGCCTTAG AAAATCGAGA 10020  
 15 TTATCAACAA TTATGTCAAA GTTTGTCTAA TCGATTAGAG CCAATTTCTG TTTCAAAACA 10080  
 CCCACAAATC GATAAATTAA AAAATAATAT GTTGAAAAGT GGTGCAGATG GTGCGTTAAT 10140  
 GAGTGAAGC GGACCTACTG TGTATGGGCT AGCACGAAAA GAAAGCCAAG CAAAAAATAT 10200  
 20 TTATAATGCA GTTAACGGTT GTTGAATGA AGTGTAATTA GTTAGACTAT TAGGATAGAA 10260  
 GGGTTGAAAA GATGAGATAT AAACGAAGCG AGAGAATTGT TTTTATGACG CAATATTTGA 10320  
 TG 10322

25 (2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5614 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

GATTGATTAA ATGTTTTAAT CCACTTCAAT GCCTTCGATA AACTCTACAA TCGCGCTATT 60  
 CATATAATTA TTCGATTTCA TTTGTTGAGC ATATGTCTCA TTAAATCCAG ACATAACTTT 120  
 40 TTTAAAWGCG AAAATTGAAA TTGGTATCGT TACTAATAAG GCACTAGCCA TACGCCAATC 180  
 AATGAGCATT ATGTATAAAA AGATAGCAGC TGACAAAAGT AAGTTTCCTA TAACTTCAGG 240  
 AATCATATGT GCTAAAGGTA ATTCTATTGT TTCAACCTTA TCGACAAATA TATTTTTTAA 300  
 45 TTCACCTATT TTCTTAGATT CCACTACGCC TAAAGGGAGA CGCATTAAAT TTTGAGCTAA 360  
 TTTTTTACGA ATTTACAGATA AAATTTTATA TGCCGTAATA TGTGATAGCA TCGTTGACGC 420  
 50 TCCAAAACAA CACACTTGTG AAATATAAGC GATTAAAGCA ATAAAGATAT AAACCATAAT 480  
 CGAATTAATC GTATATGTAT TGTTAATCAT CATTAAAATA ATTTTAAATA CTGCCCCAATA 540  
 AGGAACTAAT CCAGAAAAGA CACTGATGAT AGACAACAAA ATTGATAACA TAATTTTCCA 600

55

	ATATGTA ACT CCTkTCAATT AATAATCTAA ATTAAGCCGC TTATATTATT TATTTCACTG	720
	GATGATATAC ATAATATAAA TTTGTTATTT GTTAAAAATT AATAC TTATT ACAAGTACAT	780
5	CATATATTAG TTGATAACGA TTATCAATGT CGCGTGGATT TGTGACACAT TTCTTTTAAA	840
	AATTCACAAG GTTATGGGGC AGAAATGATA AAGAGCCACT AATGATTTAT TATGTAGTGG	900
10	TTCTGGGAGT GGGACAGAAA TGATATTTTC ACAAATTTA TTTCGTCGTC CCACCCCAAC	960
	TTGCATTGTC TCTAGAAATT GGAATCCAA TTTCTCTTTG TTGGGTCCCT GAATATAGCC	1020
	TTGTAGAGTC TAGTACATTG ATTTGTATCC CAATGTCCCT ATAATTGATT ATTCGCTTTA	1080
15	TCTAATGATC CTATGACTCA ACTATTAAAT CATTTTTTCGA AATAC TTAAAT TCTAATATAA	1140
	TTAAATTCAT TTATTGTAAT ATTGCAAAAA TACATTGCAC ACCTTGTTCA TCAATGCTAT	1200
	AATTAATTAC ATAATAAATT GAACATCTAA ATACACCAA TCCCCTCACT ACTGCCATAG	1260
20	TGAGGGGATT TATTTAGGTG TTGGTTATTT GTCACCTTTT TTATTGTTGC GCGTTCGTAA	1320
	CCAATGTGCA AAAAACGCAA CAAGACAGCC GCTTATAGCT GAAGTCATGA TGTTAATTAA	1380
	TAAATTGAAC ATCCGTCATA CACCTCCTCT CTGCGTTAAA GTAACGCCCG AGATGTTAGG	1440
25	CGACCATCAT ATTATATCAT TTATTTATTA TATTTACGC AATATTAAGG CTTAAGTAAA	1500
	GTTTTTTTTA GTGGTTTACG C TACTTTAAT TGCTATCTTT TAAAATCCAT TTAGATAATA	1560
30	TAAATGTGAT GGGTATCGTA ATAATTAAAC CAGCAAATGG TGCAATTTCT GCTGGCAAAT	1620
	TTAGCCAGGA TACAAATACA TATAATAAAA CTGTTTGTA GCTTACGTTG ACAATCTGCG	1680
	TAATTGAAA ACTAATGAAT TTTCTCCAAG TAGGTTTTAC CCTGTAAACA AAATAACAAT	1740
35	TCAAATAATA TGAAATCACA AAAGCGACTA GAAATCCGGT AATATGACTA ATCATATATT	1800
	CAATGTGTAA TAATTTTAAAC AGCAATAAAT AGACAACATA ATAATTTAAC GTATTAATGC	1860
	CGCCACAAT GATAAATTTT AAAATTTT CAG CATGCGTTTG TGTTAGTTTC ATATGTGTAC	1920
40	TCCTCAACAT CAAAATATAT GCATAACTAC GTTCTCGAAC ATACTCGAAT ATGCGAGCCA	1980
	ATCCGCTTCA CTTCAAATAT GCTTATTTCA ATCTTTATAC CCTTTCACAG CAAATTTAGT	2040
45	CTCTTTCCCC TCATCCTTAT ACGCCATTAT AATGTAAC TG ATTTATCGCG TGA CTCTTA	2100
	GCACTATAGA GATTACTTTA GTTCACTAGT AATTTTATAT ACAATAAGAG CGACAACAGT	2160
	AATGAGAGGA TGTCTACTAT GCAATTACAA AAAATTGTCA TCGCTCCTGA CTCATTTAAG	2220
50	GAAAGTATGA CCGCACAGCA AGTTGGCAAT ATTATAAAAC AGGCTTTTAC TAATGTTTAT	2280
	GGGAATACCC TTCATTATGA TATCATTCCG ATGGCTGATG GTGGTGAAGG TACCACAGAT	2340
55	GCTTTAATGC ATGCAACAGG TGCCACTAAG TATACAGTCA TCGTTAATGA CCCTTTAATG	2400

	GCGGCAGCGT CAGGTTTGGG TTTATTAGAA AAAGAGGAAC GTAATCCTTT ATACACATCA	2520
	TCATATGGTA CCGGTGAACT AATTAAAGAT GCATTAAATC ATGGTGCTAA GACCATTATT	2580
5	TTAGGGATTG GTGGCAGTGC AACAAATGAT GGTGGTACAG GTATGCTAAG TGCACTAGGC	2640
	GTAAAGTTTA CTGATGTAAA CGGGGACTTA TTACAAATGA ATGGTGCTAA TCTTGCTCAC	2700
	ATTGCACAAA TCGATATAAC CAATCTAGAT TCGCGATTAA AAGAGGTGAC CTTTAAAGTG	2760
10	GCCTGTGATG TTTCAAATCC TTTATTGGGT GAAAATGGTG CTACCTATAT TTATGGTCCT	2820
	CAAAAAGGCG CTGATGCAAA GATGATACCA AAGTTGGATT TCGCAATGTC GCATTATCAT	2880
	GATAAGATAA AAATGTGCAC AGGAAAGTCC GTTAATCAAA TACCAGGTTC TGGTGCAGCT	2940
15	GGCGGTATGG GCGCAGCATT ATTAGCGTTT TGTGAGACAA CTTTAACAAA AGGTATTGAT	3000
	GTCGTCTTTG ACATTACAGA TTTTCATCAA AGAATTAAAG ATGCAGACCT CGTTATTACT	3060
20	GGAGAAGGAC GCATGGATTA TCAGACCATC TTTGGTAAAA CACCCGTAGG CGTTGCGTTA	3120
	GCTGCAAAAC AATATCATAT TCCTGTCATC GCGATTTGTG GCAGTCTAGG CGAAAATTAT	3180
	CAACATGTTT ACGATTTTCGG TATTGATAGT GCCTATTCTA TAATCTCTTC ACCTAGCACT	3240
25	TTAGAAGATG TCCTACAAA TAGCGAACAA AATTTATTAA ACACTGCAAC TGACATTGCT	3300
	CGTATTCTGA AATTACAATA ATGTCAAAGT AAATCATCAG CTTTATTATT TGCAGTTAAA	3360
	ACTTGAATGA GGTGAAACCC ATGAAAAGAA CTGATAAATA CCGTGATTCA TATCAATACG	3420
30	ACAATCAAAA CCAAATCAT CGTCGTCAAT CTGAAGACGC ATCGTATAGA CAACAATATG	3480
	CTAAAGGCGA TCCTGAAGAA CACCCGGAAC GATACTATAA TGGTAGAGAT TATCGAAGAG	3540
	AACAAATTCT TGAAGAAGAA AACGAGAAAT CCCGCCGTTT AAAAAAATGG TTATATATCA	3600
35	TTATTGCCAT TCTCTAATT ATTGTCGCTA TTTTGTGAC ACGCGCCTTA CTTAACAATG	3660
	ATAGCGATAA AGTTAGTAAT GACCCTAAAG TCTCTCAAAA TTATAAAAAA CAAGTTGAAA	3720
40	ATCAAGACGG CCAAATTAAC CAGCAAGTAG ATAATGCTAA AGAAAATATT AAAAAACAAC	3780
	AAAAAACTGA TGACATTATT AAAAATTTAC AAAATCAAAT CGACAACCTG AAGCAGCAAG	3840
	AACAAAACAA AGCTGATTCT AAGCTAACTC AATTTTATCA AGACCAAATC AACAAATTGA	3900
45	CAGAGGCAAA TAATGCACTT AAAAACAATG CAAGCCAAGG TAAAATTGAA AGCATGTTAA	3960
	ATGATATTAA TACAAAATTC GACAGTATTA AATCTAAATT AGAAAGCTTA TTTAAAGATG	4020
	ACAATGGTGG CGCTAATTAA TTATTACACC TGCTTTGATG ATAAACATTA ATTCCCTATA	4080
50	CTTTATCTGT ATCACTACGT TATTCGTGAT GATGCATTAA GAGTATAGGG ATTTTTTATA	4140
	TAAACTTGTA TTCTAACTAC ATACAAATAC ACACAAAACG TATATAATTT ATATAATTAT	4200
55		

EP 0 786 519 A2

	TTATTGCTAA TTACGTTAGG CGTCATGACC GCTTTTGGCC CACTAACTAT AGATATGTAC	4320
	GTACCATCAT TACCTAAAGT GCAAGGTGAT TTTGGTTCTA CTACATCAGA AATTCAATTA	4380
5	ACATTATCAT TCACAATGAT TGGTCTTGCA CTAGGCCAAT TTATCTTTGG ACCTTTATCC	4440
	GATGCTTTTG GTCGCAAACG GATTGCTGTA TCCATTTTGA TCATTTTCAT TTTGGTATCA	4500
	GGTTTGCTA TGTGTGTTGA TCAATTGCCA TTATTCTTAA CTTTACGATT TATTCAAGGT	4560
10	TTAACTGGTG GTGGCGTCAT CGTGATTGCA AAAGCCTCTG CTGGTGATAA ATTTAGTGGC	4620
	AACGCACTCG CTAAATTTTT AGCATCTTTA ATGGTAGTTA ATGGCATCAT CACTATTCTT	4680
15	GCACCATTAG CCGGTGGATT AGCTTTATCC GTAGCAACAT GGCCTTCTAT TTTCACAATT	4740
	TTAACTATTG TGGCACTCAT CATTTTAATT GGCCTCGCTT CTCAATTACC TAAAACATCT	4800
	AAAGATGAAT TAAAGCAGGT GAATTTTAGT AGCGTCATTA AAGATTTTGG AAGTCTTTTG	4860
20	AAAAAACCAG CATTTATTAT TCCAATGCTA TTACAAGGWT TAACTTATGT AATGCTATTT	4920
	AGTTATTCAT CTGCATCGCC ATTTATTACT CAAAAATTGT ATAATATGAC ACCCCAACAA	4980
	TTTAGTATCA TGTTTGCTGT TAACGGTGTA GGTTTAATCA TTGTCAGTCA AGTCGTTGCT	5040
25	TTATTAGTAG AAAAATTACA TCGCCACATA TTATTAATCA TTTTAACTAT TATACAAGTG	5100
	GTAGGTGTTG CTTTAATTAT CCTGACACTT ACATTCCATT TACCACTTTG GGTCTTACTC	5160
	ATCGCATTCT TCTTAAATGT GTGTCCTGTG ACGTCAATTG GACCGCTTGG TTTCACAATG	5220
30	GCTATGGAAG AACGAACAGG TGGCAGTGGT AACGCATCAA GTTTACTTGG CTTATTCCAA	5280
	TTTATCTTAG GTGGCGCTGT TGCACCATTA GTTGGCTTAA AAGGCGAATT TAATACATCA	5340
35	CCATATATGA TTATTATCTT CATTACAGCC ATTCTATTAG TCAGTCTACA AATCATTTAC	5400
	TTTAAAATGA TTA AAAAGCA ACATGTCGCA TAACACTTCA ACATAATTAG AACCTAGCA	5460
	AAGATATCTA TCTTTGTCAG GGTCTCTCTT TATGAATTAT GAGATCGAAT CTTCAACTAA	5520
40	AATTACGCCT TCATAGCAAG GACATTTCTA TTCAATCACC CTTTAACAGG CATCCAAATT	5580
	TcTGTAATAT ATTTTTCAC TGTAGTATCA CCAT	5614

(2) INFORMATION FOR SEQ ID NO: 100:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9179 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:



	AAAGACAATG ATATGAAGTA TATGGATATC ACAGAAaAAG TGCCAATGTC GGAATCTGAA	120
	GTTAACCAAT TGCTAAAAGG TAAGGGGATT TTAGAAAATC GAGGGAAAGT TTTTCTAGAA	180
5	GCTCAAGAAA AATATGAGGT TAATGTCATT TATCTTGTTA GCCATGCATT AGTAGAAACA	240
	GGTAACGGCA AATCAGAATT AGCAAAAGGC ATTAAAGATG GGAAAAAACG CTATTACAAC	300
	TTTTTTGGTA TAGGAGCATT CGATAGTAGT GCTGTTTCGTA GTGGGAAAAG TTATGCTGAA	360
10	AAGGAACAAT GGACATCACC AGATAAGGCG ATTATTGGTG GTGCAAAGTT CATTTCGTAAT	420
	GAATATTTTG AAAACAATCA ACTGAATTTA TATCAAATGC GATGGAATCC AGAAAAATCCT	480
15	GCGCAACATC AATATGCGAG TGACATTCGC TGGGCAGATA AAATTGCCAA ATTAATGGAT	540
	AAATCCTATA AGCAGTTTGG TATAAAGAAA GATGATATTA GACAAACATA TTATAAATAA	600
	GACATCGGTG CTTAAAGGAG CTGGAACAAT TTATTGTTTC GAGCTCCTTT AGCGCATTCT	660
20	GAGTGTGTTA GTTAAATGGA TTTTAACCTA ACAAAAAACG CTATATAGCA TCAAATATGC	720
	TATATCCCAC ATCATTGTTA CAAATGTACA TGATGTAAAT GAATATTGCT GTCTAAATGT	780
	GCATGTAATA TACAATGGTG CAGATAATAC ACTTAAGTCC TTAAAAATGA AACGTTAgTT	840
25	CCAAGAGTCA TTTTAAACA ATAGTGCATG TGATAAAATA GAAAAGAATG AAAAATATAG	900
	AGGTGACAAT ATGAAGATAG CAATTATAGG TGCAGGCATC GGTGGATTAA CAGCTGCTGC	960
	ATTATTACAA GAACAAGGTC ATACTATTAA AGTCTTTGAA AAAAATGAGT CAGTTAAAGA	1020
30	AATTGGCGCT GGGATTGGTA TCGGAGATAA TGTGCTTAAA AAAGTAGGTA ATCATGACTT	1080
	AGCTAAAGGT ATTAAAAATG CTGGGCAAAT CTTATCTACA ATGACAGTGT TAGATGACAA	1140
35	AGATCGCCTG TTAACACTG TTAAATTAAA AAGTAATACA TTGAATGTGA CGTTACCACG	1200
	CCAAACATTA ATTGACATTA TTAAATCTTA TGTAAAAGAT GACGCAATAT TTACAAATCA	1260
	TGAAGTCACG CATATAGATA ATGAGACAGA TAAAGTTACC ATACATTTTCG CGGAACAAGA	1320
40	AAGTGAAGCA TTTGATTAT GTATTGGTGC TGATGGAATT CATTCTAAAG TGAGACAATC	1380
	TGTAAATGCT GACAGTAAAG TATTATATCA AGGGTATACA TGCTTTAGAG GTTTAATTGA	1440
	TGATATTGAT TTAAAGCATC CGGATTGTGC AAAAGAATAC TGGGGaAGAA AAGGaAGAGT	1500
45	AGGTATTGTT CCGTTATTAA ATAATCAAGC ATATTGGTTC ATTACAATTA ACTCGAAGGA	1560
	AAACAATCAT AAATATAGTT CGTTTGGTAA ACCTCATTTG CAAGCATACT TTAATCACTA	1620
	TCCAAATGAA GTTAGAGAGA TCTTAGACAA ACAAAGTGAA ACAGGTATCT TATTGCATAA	1680
50	TATTTATGAT TTGAAACCAC TCAAATCTTT TGTTTATGGT CGTACTATTT TACTAGGAGA	1740
	TGCAGCACAT GCGACAACGC CTAATATGGG GCAAGGTGCT GGACAAGCAA TGGAAGATGC	1800
55		

	TAAAATACGT	GTCAAACATA	CTGCAAAAGT	AATTAAGCGT	TCTAGAAAAA	TCGGTAAAAT	1920
	TGCCCCAATAT	CGTAGTCGTT	TATTTGTTGC	AGTTAGAAAT	CGTATTATGA	AAATGATGCC	1980
5	AAATGCATTA	GCAGCTGGAC	AAACTAAATT	CTTATATAAA	TCGAAAGAAA	AATAATACAA	2040
	CAATATGAAA	ACCCCCGTAT	GTTGAAACGA	GAGCTCAACA	TATGGGGGTT	CTTGTTTTTA	2100
	TAATGTTATT	ATAATAAATT	CAATTATTAG	TTAACGACAA	ATTGTGGTTT	CTCACCTTGA	2160
10	ACGGCACTAA	TTGCAGCATT	AGCAACAATT	TTAGACATCA	TGTCACGTGC	TTCAAATGTA	2220
	GCATTACCAA	TATGCGGTGT	TAATACTACA	TTATTAAGTG	ATTTTAAGTC	ATCGGTAATA	2280
15	TCTGGTTCAA	ATTCATATAC	ATCAAGTGCA	GCACCTTCAA	TTTCATTATC	TTTCAATGCT	2340
	TGCACTAGTG	CTTGTTCTGT	CACGATTGGA	CCACGAGAGG	CATTGATTAA	ATACGCCGTA	2400
	GATTCATCA	TTTTAAATTG	TTCTGTATCA	ATTAAATGAT	GCATTTTAGG	ATTATAAGCA	2460
20	GCGTTGATAG	TGATAAAATC	TGCATTCTTT	AATAGTGTAT	CTAAATCTAC	ATATTTTGCA	2520
	CCGATTTCTC	GTTCTTTTTT	TTCTTTGCGA	TTAGGTCCAG	TGTATAGCAC	ATCCATGTCA	2580
	AATGCTCTTG	CACGACGAGC	TACTGCACTA	CCAATTTTAC	CTAAACCGAT	AATGCCGATT	2640
25	GTTTTCCCAG	ATACTTCTCT	ACCTCTGAAA	AATAAAGGTG	CCCATCCATC	AAATCCAGTT	2700
	GTACGTGATA	ATTGGTCCCC	TTCAACAATA	CGACGCGCTA	CTGCAAGTAC	TAATCCAATT	2760
	GTTAAATCAG	CAGTCGCGTT	TGTTGATGCT	TTAGGTGTGT	TTGTAACATC	TATACTTTTT	2820
30	TCTCGGGCAT	ACTCGATATC	AATATTATTA	AAACCAGCGC	CATAGTTGGC	AATGATTTTT	2880
	AAGTCTTTAC	CAGCATCGAT	AACATCTTTA	TCAACGTTTG	TAGATAATAA	ACTAATTAAG	2940
35	GCAGTCGCGT	TTTAAACACC	TTTAATTAAA	GTGTCTTTAT	CGACTAATCC	TTTACCTTCA	3000
	TACATTTCAA	CTTCAAAATG	TTCTTGTAAG	AGTTTTTAAAC	CTACTTCTGG	TATtGCACCA	3060
	gCAACATAAm	CTTTTtCCAT	AAAAGAtCAC	TCCTTTTATC	TTAGTATAGT	AGAAGATTAG	3120
40	ACAGTATACA	ACTATGTCAT	GATGTCTTGT	GTATCAATGA	TGTAAGCGCG	TACTTTTGAT	3180
	GGAGGCGATA	TAACCTTAGGC	ACTGTAGAAC	TATGAATATT	GTAATGTGGA	AAAACGGAT	3240
	CAATTAAATT	AGATAACGTA	GTTTTTAAAGT	TAATAGTATT	AGAAAAAATT	AATATTTTGA	3300
45	ATATGGGAGG	AAATATAAAT	AAGTAGGTGG	CAACGAAAAA	TAGCAAAAAA	AGAGCTTCTC	3360
	CTATAAAGGA	AAGCTCAAAG	TTTTTTGATG	ACATATGTAC	TAGAATTAAG	TTTCAAGACA	3420
	ATATGTATCA	TCGTGTTTAT	ATTAAATATG	GATGTAGTTG	TAGTTACCTG	CTTCACTTGC	3480
50	AGAAATAGTT	CTAGAACTTA	CTGAGAAAGG	TCCGCCACTA	TAATTCATTT	CTGAAATTGT	3540
	AACTGAACCA	TCACTGTTTA	CACTTTCTAC	ATATGCAACG	TGACCAAATG	GTCCTTCAGA	3600

55

EP 0 786 519 A2

	AGCAGCAGCC CAATTATTAG CATTTCCTCCCA AGTAGAACCG ATTTCTCCGC CAACTTTATC	3720
	ATATACATAC CAAGTACATT GTCCTGCAGT GTATAAGTTA CCAGAATGTG AAATTGATGA	3780
5	TGTAGTTGTC GTAGTTGTCG TAGTCGTTGT AGTTTGAGTC GTGTTGTAGT TATAGTTGTT	3840
	GTAATTTGTA TAATTTTCAG CAGCATCTGC ATGATGTGCT TGACCTACTA ATGCTGTGCC	3900
	GATTCCTGCT GTTAACGTAG TTGCTGTTAC TAATTTTTTC ATGAATAAAG TCCTCCAAAG	3960
10	TTCTATATCT TTTTTTATAA ATAAAACGTA GCGACTGTTT TATTCTCACA TCTCGAATTG	4020
	ATGACAATAG TTACTTTAAC AAAATTAATG CTTCTTGTGG GGAATGTTAT TGATTTGTAA	4080
15	AAGAATAAAA AAACCTTGAC TAATTTTGTA ATAAAAATTA GTCAAAGTTA CAATGAGATT	4140
	AACAGATAAT TAATAGGAAA TATTTATTG TAATATGTTT AAATAAATCG AATTGTTAAA	4200
	GGTATTATAT ATTCTTGGCC ATTATAATAT TTGACACACG CAATAATTGT GAATACAAAA	4260
20	GATAATATTG AGAAAGCGAA TATGGATAAA ATACCGATAA ACGTAATGAT GAAACCTATA	4320
	ATAATAATGA AATCAATATC TGTAGCAATT AGGAAAACGC CTATTAAAGT GATAACGACT	4380
	AAAACGATAG ACCAAATAAT ATAAGAAATC GTATAGTTAA GATAATTTTT TCCAGCACGA	4440
25	TCAACTAGTT TCGATTCATC TTTTTTCAAT AACCATATTA TCAGTGGACC AATAATAGAT	4500
	GTGAATAAAC TTAATAAATA GATAAGCATC GCCATAATGT TCTCATCATT GGATTTGCGA	4560
	TTCGGTTGAT GATTTGTTAC GTCGTTTATT TCAGTTGTCA TATTAGACAC TCCTTTGAAA	4620
30	ATTGTAATAT TATCTTTAAC TATAACAAAA TATAATCAAA AATAAACATG TTTATTAAAC	4680
	AATTATTAAA AATAAAAATA ATTGGTGGAC GTCGGCGTTT AAATAGGTTA ATTTAAGGTT	4740
35	ATATATACTT AACATTTATA ATGATGCGTA ATGAATTCGC ATCATTTTTA TATTGTCTTA	4800
	CGTATAATTT GTTTTTAATT TTAACCAAAG ATAGAAAGAG GGTTGTTTAT GAAAATAGCA	4860
	ATTGTAGGAT CAGGAAATGG CGCAGTTACG GCAGCAGTAG ATATGGTGAG CAAAGGCCAC	4920
40	GATGTTAAAT TATATTGTCG TAATCAATCT ATAAGTAAGT TTCAAAACGC AATCGAAAAG	4980
	GGCGGATTTG ATTTTAATAA TGAAGGTGAT GAACGTTTCG TAAAATTCAC TGATATTAGT	5040
	GATGATATGG AATATGTTTT AAAAGATGCT GAAATTGTTC AAGTGATTAT TCCATCTTCA	5100
45	TACATAGAGT ATTATGCTGA TGTAATGGCA GAGCATGTAA CTGATAATCA GTTGATATTC	5160
	TTCAACATGG CTGCAGCAAT GGGGTCAATT CGTTTTATGA ATGTTTTAGA AGATAGACAT	5220
	ATTGAAACAA AACCACAAC AGCGGAAGCT AATACGTTGA CGTATGGTAC GCGTGTGAT	5280
50	TTTGAAAATG CAGCAGTTGA TTTATCTCTA AATGTACGTC GTATCTTCTT TTCAACATAT	5340
	GATAGAAGCT GTCTAAATGA TTGTTATGAC AAAGTTTCAA GTATTTATGA TCATTTAGTA	5400
55		

	CCAACATTAT TGAATGTCGG TCGCATTGAT TATGCTGGCG AGTTCGCTTT ATATAAAGAA	5520
	GGAATTACTA AACATACAGT TAGATTACTT CATGCAATCG AATTAGAACG TTTGAATTTA	5580
5	GGCCGTAGAT TAGGTTTTGA ATTATCAACA GCTAAAGAAT CACGTATTGA ACGTGGTTAT	5640
	TTAGAACGTG ATAAAGAAGA TGAACCATT AATCGTTTGT TTAATACAAG CCCAGTATTT	5700
	TCACAAATTC CAGGACCAAA TCATGTAGAA AGCAGATATT TAACTGAAGA TATTGCATAT	5760
10	GGTTTAGTAC TATGGTCAAG CTTAGGTCGT GTTATTGATG TACCGACACC AAATATAGAT	5820
	GCAGTAATTG TAATTGCATC AACCATTTTA GAGAGAGACT TCTTTGAGGA AGGCTTAACA	5880
15	GTTGAAGAAA TTGGTTTGA TAAGCTTGAT TTAGAAAAAT ATTTAAATA AATGATGGCT	5940
	TGAAGATAGA AAAGGATATA GCATTATGCA AAAGCAATAA ATTGAAGAAA AGAGGTTTCT	6000
	CATCAATAAG CGnAGGGGAC GATAGATGAT GAAAAGAAAA CCCACCTTTT TAGAATCAAT	6060
20	TTCGACAATG ATTGTAATGG TTATTGTTGT TGTAACAGGC TTTGTGTTTT TTGATATTCC	6120
	AATTCAAGTA TTATTAATTA TTGCCTCAGC ATATGCCACA TGGATTGCAA AACGTGTAGG	6180
	CTTAACATGG CAAGATTTAG AAAAAGGCAT TGCAGAACGT TTAAATACTG CAATGCCTGC	6240
25	AATTTTAATT ATACTAGCGG TAGGAATTAT AGTAGGCAGT TGGATGTTTT CTGGCACAGT	6300
	GCCAGCCTTG ATTTATTATG GCTTAGATTT ATTGAATCCA AGCTATTTTT TAATATCAGC	6360
	CTTTTTTATA AGTGCTGTTA CATCTGTAGC AACTGGTACA GCATGGGGCT CTGCATCAAC	6420
30	TGCAGGGATT GCACTTATTT CTATTGGTAA TCAATTGGGG ATTCTCCAG GGATGGCAGC	6480
	GGGTGCTATT ATAGCAGGGG CTGTGTTTGG CGATAAAATG TCACCATTAT CAGATACAAC	6540
35	TAATTTAGCG GCGCTTGTTA CTAAAGTTAA TATATTTAAA CATATACATT CGATGATGTG	6600
	GACGACGATA CCTGCATCAA TCATAGGTTT ATTAGTATGG TTTATTGCTG GATTTCAATT	6660
	TAAAGGGCAT TCAATGATA AACAGATTCA AACTTTGTGA TCAGAGCTTG CACAGATTTA	6720
40	TCAAATTAAC ATATGGGTCT GGGTTCCCTT AATTGTGATC ATTGTTTGTT TGCTATTTAA	6780
	AATGGCTACA GTGCCAGCTA TGCTAATATC AAGCTTTTCT GCCATTATAG TGGGGACTTT	6840
	TAATCATCAT TTCAAATGA CAGATGGTTT CAAAGCAACA TTTAGTGGTT TTAACGAATC	6900
45	AATGATACAT CAGTCTCATA TTCATCCAG TGTGAAAAGC TTGTTAGAAC AGGGTGGTAT	6960
	GATGAGTATG ACCCAAATAT TAGTAACGAT ATTTTGCAGA TATGCATTTG CAGGTATTGT	7020
	AGAAAAAGCA GGATGTTTAG AAGTCTTATT AACTACTATT TCTAAAGGCA TCCATTCTGT	7080
50	AGGAAGTTA ATATGTATTA CTGTTATTTG TTGTATTGCG CTTGTATTCTG CTGCAGGTGT	7140
	TGCTTCGATT GTAATTATTA TGGTCGGTGT GTTAATGAAA GATTTGTTCG AAAAATACCA	7200
55		

EP 0 786 519 A2

	AATACCATGG	GGAACATCAG	GTATTTACTA	TACGAATCAA	CTTCATGTCT	CTGTTGAAGA	7320
	ATTTTTTCATA	TGGACAGTAC	CATGTTATTT	ATGCGCAATT	ATAGCAATTA	TCTATGGTTT	7380
5	TACAGGGATA	GGTATTAAAA	AGTCATCGAA	TTCACGTTTA	ACTTAATGTG	AGCGTGGAAT	7440
	ATATATAATA	TGTTGAAACA	CTTTAATCAT	TTATAATTGT	AGCGGTTATA	ATTTGAAAAG	7500
	GTTTTAACTT	AGAATAAATA	TCCTCTATGC	ATATACTGAA	TATGTTTTGT	AGCGGAACAT	7560
10	GTTGATATAT	GTAATGTAAG	TTTTATGTCA	TGATTTGTAA	TGACTAAATT	AATTGAGAAT	7620
	TTGAAGGCAA	GTATATTGT	AAGTACTTTA	ACTAAAAATT	TATCAATGTA	TAGCCGATTT	7680
	GACATGCCTA	AATTTGGGTG	TGTCAATGGC	TGTATGTTGT	TTATTCTTTA	TTACAGAGTG	7740
15	AATCGGATTG	GTGAAAATCG	AAATTTTGAG	ATTTTACCA	ATTCGATTTT	TTTCATAGAA	7800
	ATTAATAAAG	CCAACAAGGC	TCTTGAAACC	TTGTTGGCGT	AAACATAGCC	ATCACTAATT	7860
20	AGTGAATGAA	GTTATAACCA	GCAGCTTGGC	TAGCTGAGAT	TGTACGTGAA	GTTACAACAC	7920
	CTGGGCCATA	ACCATAGTTC	ATTTCTGAAA	CTCTTACTGA	ACCATTGCTG	TTAACACTTT	7980
	CAACGTATGC	AACGTGACCG	TATGCACCTT	GAGTTGTTTG	CATAATTGCA	CCAGCTTTTG	8040
25	GTGTATTGTT	CACTGTGTAA	CCAGCTCTTG	CAGCTGCGTT	AGCCCAGTTA	CTTGCATTGC	8100
	CCCAAGTTGA	ACCGATTTTA	CCACCTACAC	GATCAAATAC	GTAGTATGTA	CATTGACCAG	8160
	AAGTGTATAA	GTTACGTCCT	GAAGTATAAC	CACTTGAGAT	TGAACGGCCA	TTTGATGATG	8220
30	GAGCCATAGT	TGTAGTTACT	TGAACATTGT	TGCTTGAAGT	GCTGTAGCTT	GCACCTAAAC	8280
	CACCAGTACG	GTAGCTGTTT	GTGTGTAAAC	TATTATAGTT	ATTGTAGTTA	TATGATTGAT	8340
	TATTATTTGA	GTAGTTGTTG	TAACGGCTGT	AGTTATTGTA	GCTATAACCG	TTGTTGTAAT	8400
35	TGTTATAGTT	ATTGTAACCA	TTGTAGTAGT	AATAGCTGTA	GTAGCCATTA	TCTTGGTTTA	8460
	ATTGACTTGG	ATGCCAGTTA	CCTTTCCATG	TGTAATGGTA	GTTACCTTGT	GCATCAATAG	8520
40	TGTAAGTATA	GCTATATGAT	GTTGGGTCGT	TTGGATTATA	ACCGTAGTTA	TCTTGCTCAG	8580
	AAGCATGAGC	TTGATTTCTT	GATGCAATTG	CGATTGTAGC	GAATCCTGCA	GTTGCGATAG	8640
	TAGCTGTAGC	GATTTTCTTC	ATTTTAAAAA	TATCCTCCTA	AAAATTTTAA	ATCTAAAATA	8700
45	TTTTTCGTAAT	GTCCGTGTGA	CAAATAAAT	GTTATAAGTT	ATCTCTCGTA	ATTAAACGAC	8760
	AAGAAAGACT	ATAACAGAAA	TTAGCGTCCT	TGTGTGCTTT	GTTAACGTTT	TGTAATTTTT	8820
	TGCTAATATC	TTGACACAAT	AGAATTTTAA	AAGTATAGAA	ATTTGCATTT	TGCAAACTT	8880
50	ATAACTACGG	CATTCTTTGT	GAAAAC TGAA	TGTTTCGAAA	ATAAGTCTGT	TACAAATTTG	8940
	TAATATTACT	GAAAATTCTA	AATGTATATT	TTGTGCATAA	TATAGGACTT	TTAATCAGAA	9000
55							

GGATGAAAAT GTATATTTAA TGGATAAAAT ATCCTAATTT AGCATAAAAA AATGTTTTAA 9120  
TAAAAGTATT ATTTGATATA ATCGATTAT GTTTTGTTAC TGCTAAAAAA CATGTGGCG 9179

5 (2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1868 base pairs  
(B) TYPE: nucleic acid  
10 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

15 CCTTCAGCCA TTTGACTTCG ACATGAGTTG CCTGTACATA TAAAATAAAT TGTTTTTTTA 60  
GTCATAACAA TCTCCTAATT AATTAAAATA TGATAAGTGT TAGATACAAC CCTATGAGGG 120  
20 TTATAAATAG TACTGGAATT GTAATGATGA TACCAGTTTT AAAGTATGTG CCCCAAGAAA 180  
TCTTAACATC TTTTGTGTT AAGACGTGTA ACCACAGTAA TGTAGCTAAA GAGCCTATCG 240  
GTGTAATTTT TGGACCTAAA TCAGAACCGA TAACATTCGC ATAAATTAGG CCTTCTTTTA 300  
25 ACATGCCATG GACATTTGAT TGACCAATAG CAATCGCATC TATTAAAACT GTAGGCATAT 360  
TATTCATTAT TGATGATAAA AACGCTGAAA TGAAGCCCAT TCCCAAATA GTGCTAAATA 420  
GACCGTAATT GGAAATATAT TCTAATATTT TAGCCAATAT TAAAGTAATG CCAGCATTTC 480  
30 TTAAGCCGAA TACGACGATA TACATACCAA TTGAAAATAA TACTATATTC CAAGGTGCGC 540  
CCTTAATGAC TTGCTTAATA TTTACAGCAT TTGATTTACG AGCCAACATT AGAAAAATAA 600  
AAGCAATGAT TCCAGTGAAA ATTGATACCG GAATTTTAGT AAATTTACTG ATTAGATAGC 660  
35 CGAAAAGTAA TATAACTAGA ACAATCCaTG AAATTTTAAA TAGCTTTAAA TCATTAATGG 720  
CATCFTTAGG ATGCTTTATA TTATTATCAT CAAACGTTTT AGGTATCGCT TTTCTAAAAT 780  
40 ATAACCACAA TACTATAATA CTTGTAAAA GCGAGAATAA ATTAGGTATA ATCATTCTAC 840  
TAAAATATCG AACGAATCCT ACATGAAAAT AATCAGCAGA TATAATATTC ACTAGATTGC 900  
TCACGATTAA AGGTAAAGAA GTTGTGTCAG CTATAAAACC ACTCGCAATA ATnAAAGGGA 960  
45 ATATGGCCCG CTTACTAAAA CCTATATTTT TAACCATCGC TAATACAATA GGCGTTAAGA 1020  
TTAAcGTGCG CCATCATTG CGAAAAATGC AGCAACAATG GCACCCAATA ATATGATATA 1080  
AACGAACATT TTAAACCAT TGCCTTTTGA AGCATGAAGC ATGTGAATAG CTGACCATTTC 1140  
50 GAATAATCCA ACTTTATCTA ATATTAATGA AATAAGAATG ACTGAGACAA AAGTCAAAGT 1200  
AGCATTCCAA ACAATACCTG TTACTTCGAA AACATCGGAA AAACCTTACAA CACCAGTAAT 1260

55

TAATACAAAT AATAAAGTTA CTAGAAAAAT GAGTGTGCT AAAGTTGTCA TCATTAGCAT 1380  
 TCACCAGTCT TAAGGTTATG ACAAATACAT CGTTGGTTAG AGGTATGAAC CTTAGACAAG 1440  
 5 TTATTAATTA CGGACTCAA AATATTATGA TTgAGCTGGT ATAAATGTTT ATTTCCGATT 1500  
 TTTCGTGTCG TAACTAAGTT GGTTTTACT AATGCTTTCA TATGrTAGCT AAGTGTAGGT 1560  
 TGAGAGAATT GAAAATGTGC TAACAAATCA CAAGCGCATA ACTCTCCACA AGAAAGTAAA 1620  
 10 TCTAGTATTT CTAATCTGCT TGAATCTGAT AAACTTTTA AAAATGTTGC TAGTTCTTTA 1680  
 TACGTCATAA CATACCTCCT AGACGTTAAA TAGATTATCA TCTATATAGA TGAATGTCTA 1740  
 TGTTCTTTG GTATATTACA CGATATGACT ATGTAATTTA AATTGTTTT TAGTATTAAA 1800  
 15 AGGGTATTAA AGATAAATTA TAGATATTGA TTTTGCAAAA TATACTCTTT GTTCTGCATT 1860  
 GAAAAAGG 1868

20 (2) INFORMATION FOR SEQ ID NO: 102:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15249 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

30 ATTTATGAAA TCCATAGCA TAAACATTAT TCTTGCATCG GCTATACAAA CAGTTACCGC 60  
 AAGCAAATTT GTATATCAAC CTGGAATTGT GTTCACGTCA ATGGCaATG CCGATGATGT 120  
 GTTATCAGGC GATAGTTATT TTATGGCTGA ATTAAATCT ATTAAGCGTA TTGTTGAAAT 180  
 35 TCCAGATAAT CAAAAAATAT ACTGCTTTAT AGATGAAATT TTTAAAGGTA CCAACACAAC 240  
 TGAAcGAATT GCCGCTTCAG AATCAGTACT ATCATTTTTA CATGAAAAAT CTAACTTTAG 300  
 40 AGTTATTGCA GCAACACATG ATATTGAGTT AGCTGAACTC TTAACAACAC GTTATGAAAA 360  
 TTACCATTTC AATGAGGTAA TAGAAAATAA TAACATACAT TTTGATTACA AAATTAAGCC 420  
 TGGCAAAGCA AATACACGTA ATGCCATCGA ATTATTAAAA ATCACTTCAT TTCCAGCAAA 480  
 45 AATATATGAA CGAGCAAAG ATAATGTCCC GAAAATTTAG CATTTAACCT TAAACATAAA 540  
 AACGTCAGCT ATCACATGAC AGAAGACTAT GAACAGTTTC AATAATGTTC ATAGTAATCA 600  
 TGTTAATAAC TGACGTTTAT TTTATTCTGC AGAATACTCT TCTAAATCTA TATTGCTGTG 660  
 50 CCCATTTAAT GCTAAATCAG CAAATCGACC TTGCTGATAC AAATAGTGGC CGGCAACGCC 720  
 TATCATTGCA GCATTATCTG TGCATAATTT AGGACTTGGG ATAGTTAATT GAATGTCATT 780

55

	AACAATTAAT CGCTGAACAC CATATTCTTT ACAAGCTTGA ATAGCTTTAA ACGTGAGCAC	900
	CTCTACAACA CTGTTTGGAA AGCTCGTTGC TACGTTAGCT TCAATGATTG GaATATTTTT	960
5	TTGTCGTTGA TTGTGAAGTT GATTGATTAC GGCACTTTTC AACCCACTAA AACTAAAATC	1020
	ATAACTATCT TTATCCAACC AAACACGAGG GAATGAATAA GTATCTTCAC CTTCAGCAGC	1080
	CAACCGATCA ACTTGTGGAC CACCTGGATA ATTTAAACCA ATTGTTCTGT CCACTTTATC	1140
10	ATAAGCCTCA CCTACTGCGT CATCTCGTGT TTCACCAATG ACTTCAAATG ATAAATGATC	1200
	CTTCATATAA ACTAATTCAG TATGTCCACC TGAAACAATA AGTGCAATTA GCGGGAATGT	1260
15	TAATGGCTCT TCTATGTGAT TAGCATATAT ATGTCCTGCA ATATGATGAA CAGGAATAAG	1320
	TGGCTnATCG TAAGCAAATG CCAATGCTTT GGCTGCATTA ACACCTATTA GTAACGCACC	1380
	AATTAGTCCA GGGCCTTCTG TAACCGCTAT GGCATCAATA TCTTCTATTG ATACATCGGC	1440
20	ATCCCCTAGA GCCTCGTTTA TTGTTGCTGT TATACCTTCA ACGTGATGTC TACTTGCCAC	1500
	TTCGGGAACG ACACCGCCAA ATCGTTTATG ACTTTCAATC TGACTTAAAA CTGTATTTGA	1560
	TAAAAATATCT CTGCCATTTT TTATAACACT AACGCTTGTT TCATCACAAC TTGTTTCAAC	1620
25	AGCTAGTATT AATATATCTT TAGTCATTTA AATTCACCCA CATAACCATT GCGTCCTCAC	1680
	CTTCACCATA ATAATTTTTA CGTTTACCAC CATATTGAAA TCCTAAATTT TCATATACAT	1740
	GTTGTGCCAC TTTATTATTA ACTCTTACTT CTAAACTCAT CACATCACAA GTGTGACTTG	1800
30	CATAGTTTAT TCCGTATTTT AAAAGCATTT GACCTAAACC ATAGCCTCTA TAATTATCAT	1860
	CGATTGCAAC TGTGTGAATT TGAGCTTGAT CGATAACAAT CCATAAACCT AAATAACCAA	1920
35	TAATTGTGTT TTCAAATTCt AAGACAAAAT ATTTGCAAAA GTTATTTTGC TCTATTTTCA	1980
	GATAAAATGC GTCAATTGTC CAAGAACTGT CATTGAAACT CCGACGCTCA AGATCAAAGA	2040
	CTTGTTGGCAC ATCTTCTTTA GTCATCTCTC TAATGTTTAA TTGTTCTTTT GACTGTTGAT	2100
40	CCAATTTTCGT TCCGCCTCAG CTAATTTATG GTATTTAGGA GTAAATGTAT GTACGTCTGA	2160
	AGGTTTATCT AGCAATTGAT ACATGACTGA TGCATTTGGT AGctGCGCAA TCACTTCACC	2220
	TTGTAATTCA TCTTGTAATT TTACAGTATC TTTCCCAATA TAAATAAATG GTTGGTTTAA	2280
45	ATCTTCTAAA AAAGCTCGCA ATGCCTCTAT CGACATATAT TGATCTTCTA AAATAGTCAC	2340
	TAATTGACCA TTTTGCCACT GGAATATGCC TGTATAAACT GCTTGTGCTC TTGCATCAAA	2400
	CACAGGAACC AATAATTTAT CAGTATGATC GATTGTTGCT GCCAATGCCT TTAATGATGA	2460
50	AACACCATAT AATTTAACAT CTAACGCATA CGCTAATGTT TTAGCAACAG TAACACCGAT	2520
	ACGTAAGCCA GTATATGAAC CAGGACCTTC AGCAACAATA ATCGCATCTA ATTGCTGTTT	2580
55		



	TTGTTTAGAA TCCGTAGTTA TTTCAGCTAA AACTTCATCG TTTTGCATCA ATGCTACTGA	2700
	TAATGGTTGA TTCGATGTAT CAATGAGCAG CGAATTCATG GATAATTGCC TCCTTAATTT	2760
5	G TTCATAATG TTCTCCTTGC GCGAACAAC T CAATTTGTCT TGTATTTTCA GATATTGTTG	2820
	AAATGTTAAT AGATAAATGC GTCGCTGGAA GTAAATCTTT TATAAATTGA CTCCATTCAA	2880
	TAACAGTAAT TGCCTGATCT TCGAAAAATT CATCAAATCC TAAATCTTCA TCAGAATCTT	2940
10	CTAAGCGATA ACAATCCATA TGATGCAATT TTAAATTTT ACCCCTATAT GATTTAATGA	3000
	TGTTAAATGT CGGGGAATTA ATCGTACGTC TTACACCAAG AGCTTTTCCT ATAAATTGCG	3060
15	TTAACGTTGT TTACCTGCT CCTAAATCTC CGTTAAGTAA AATCAAATCA CCACTTTTCA	3120
	ATTGCTCAAC TAAAAATATA GCAAATTGAT TCATTTTCATC TAAATTATTT ATCTTTATCA	3180
	ATGTTGATTC TCCTATATTA TGCTTTTCAT TCATAAAAAT GATTATCCAT TGTTCATCG	3240
20	TATCTAACTT TATATTTAAC CTTTATATTG TAACAAATTT CAACTTAAAT TTCTTATCTT	3300
	TGAAACAGAT TATCTATTCA AAGTTAATTG TAAGAAAATT TAAAATATTT GTTGACATAC	3360
	TAAAGCAGAT ATAGTAAATT AAATTTATCA AATTTT TAGA CAATTCTAAC TATTAAAGTG	3420
25	ATATATACCA TTCACGGAAG GAGTATAATA AAATGCTTAA TCAATATACT GAACATCAAC	3480
	CGACAACCTC AAATATTATT ATTTTATTAT ACTCTTTAGG ACTCGAACGT TAGTAAATAT	3540
	TTACTAAACG CTTTAAAGTCC TATTTCTGTT TGAATGGGAC TTGTAAACGT CCCAATAATA	3600
30	TTGGGACGTT TTTTATGTT TTATCTTTCA ATTACTTATT TTTATTACTA TAAAACATGA	3660
	TTAATCATT A AATTTACGG GGAATTTAC TATGCGAaCG AgcATGATCA AAAAAGGAGA	3720
35	TCACCAAGCA CCAGCAAGAA GTCTTTTACA TGCCACGGGC GCGCTAAAAA GTCCAACTGA	3780
	TATGAACAAA CCATTTGTAG CTATTTGTAA CTCTTATATT GATATTGTTT CTGGACATGT	3840
	TCAC TTGAGA GAGCTTG CAG ATATAGCTAA AGAAGCAATT AGAGAAGCCG GTGCCATTCC	3900
40	ATTGAATTC AATACAATTG GTGTTGATGA TGGAATAGCT ATGGGACATA TCGGAATGCG	3960
	ATATTCTCTA CCATCACGTG AAATTATTGC AGATGCAGCT GAAACTGTAA TTAACGCTCA	4020
	TTGGTTTGAC GCGGTATTTT ACATTCCTAA TTGTGACAAG ATTACACCCG GTATGATTTT	4080
45	AGCAGCCATG AGGACAAACG TACCAGCTAT CTTTGTGCTCT GGTGGACCAA TGAAAGCTGG	4140
	CTTATCTGCA CATGGAAAAG CATTAACTT TTCATCAATG TTTGAAGCAG TCGGCGCATT	4200
	TAAAGAAGGA TCGATTTCTA AAGAAGAATT TTTAGATATG GAACAAAATG CCTGCCCTAC	4260
50	TTGTGGTTCA TGTGCTGGGA TGTTTACTGC AAATTCAATG AACTGTTTGA TGGAAGTTTT	4320
	AGGTCTAGCA TTACCATACA ACGGTACTGC ACTTG CAGTC AGTGATCAGC GACGAGAAAT	4380
55		

	TATCGTTACT	CGCGAAGCAA	TTGATGATGC	ATTTGCACTT	GATATGGCTA	TGGGTGGTTC	4500
	AACAAACACG	GTA CTGCATA	CGTTAGCCAT	TGCCAATGAA	GCTGGTATTG	ATTATGACTT	4560
5	AGAGCGCATT	AATGCTATTG	CCAAACGCAC	GCCATATTTA	TCAAAAATAG	CACCTAGTTC	4620
	ATCGTATTCA	ATGCATGATG	TGCATGAAGC	TGGTGGCGTC	CCAGCAATTA	TTAATGAATT	4680
	GATGAAGAAA	GATGGCACGT	TACACCCAGA	TAGAATCACA	GTTACTGGCA	AAACGTTACG	4740
10	TGAAAATAAC	GAAGGCAAAG	AAATTAAGAA	CTTTGATGTC	ATTCACCCTC	TTGATGCACC	4800
	ATATGATGCA	CAAGGCGGTT	TATCTATCTT	ATTTGGTAAT	ATCGCCCCTA	AAGGCGCAGT	4860
15	TATTAAAGTT	GGCGGCGTTG	ATCCATCTAT	CAAAACATTT	ACTGGGAAAG	CAATTGTGTT	4920
	CAATTCGCAT	GATGAAGCTG	TTGAAGCAAT	AGACAATCGT	ACCGTTCGTG	CAGGCCACGT	4980
	CGTTGTCATT	AGATATGAAG	GACCTAAAGG	TGGACCAGGT	ATGCCTGAAA	TGTTAGCACC	5040
20	TACTTCCTCT	ATTGTTGGTC	GCGGCTTAGG	TAAAGATGTT	GCATTAATTA	CTGATGGGCG	5100
	TTTTTCCGGT	GCCACAAGAG	GTATTGCAGT	TGGTCATATT	TCCCCTGAAG	CTGCATCTGG	5160
	TGGACCAATT	GCCTTAATTG	AAGATGGTGA	TGAGATTACT	ATTGATTTAA	CAAATCGTAC	5220
25	ATTAAACGTA	AACCAGCCTG	AAGATGTTCT	AGCGCGTCGC	CGAGAATCTT	TAACACCATT	5280
	TAAAGCGAAA	GTAAAAACAG	GTTATCTAGC	TCGTTATACT	GCCCTAGTAA	CTAGCGCAAA	5340
	TACAGGTGGC	GTCATGCAAG	TCCCTGAGAA	TTTAATTTAA	TTTATTTTTA	TATTGGAGAT	5400
30	GGTTAAATG	TCTAAAACTC	AACATGAAGT	AAACCAAAT	ATTGACCCTT	TAAAAATGGC	5460
	TGAATCACTT	GAACCTGAAC	AACTAAATGA	AAAACTTTA	AATGATATGC	G TTCAGGATC	5520
35	AGAAGTGCTA	GTAGAAGCTC	TACTTAAAGA	AAATGTGGAT	TATTTATTTCG	GTTATCCTGG	5580
	TGGTGCCGTA	CTACCTTTAT	ATGACACGTT	TTATGATGGT	AAAATCAAAC	ATATTTTAGC	5640
	AAGA CACGAA	CAAGGTGCTG	TTCATGCTGC	AGAAGGTTAT	GCACGTGTAT	CTGGTAAamT	5700
40	GGCGTCGTTG	TAGTTACAAG	CGGTCCaGGT	GCAACTAATG	TAATGACAGG	TATTACGGAT	5760
	GCACATTGCG	ACTCTTTACC	TCTAGTTGTA	T TCACTGGAC	AAGTTGCTAC	ACCAGGCATT	5820
	GGTAAAGATG	CATTCCAAGA	AGCGGATATT	CTATCTATGA	CTTCACCAAT	TACAAAACAA	5880
45	AATTATCAAG	TGAAACGTGT	TGAAGATATC	CCTAAAATCG	TACACGAAGC	TTTCCATGTA	5940
	GCTAATTCTG	GACGCAAAGG	TCCTGTAGTG	ATTGATTTTC	CAAAAGATAT	GGGTGTTTTA	6000
	GCTACAAATG	TGGATTTATG	CGACGAAATC	AATATTCCAG	GTTATGAAGT	TGTTACAGAA	6060
50	CCAGAAAATA	AAGACATTGA	CACTTTTCATC	TCACTTTTAA	AAGAAGCGAA	AAAGCCTGTC	6120
	GTATTAGCCG	GCGCAGGTAT	TAATCAATCA	AAATCAAATC	AATTATTAAC	ACAGTTTGTT	6180
55							

	GATACACTAT TTTTAGGTAT GGGAGGAATG CATGGTCTT ATGCTAGTAA CATGGCATT	6300
	ACTGAGTGTG ATTTACTCAT TAATTTAGGT AGCCGCTTCG ATGATAGATT AGCAAGCAAA	6360
5	CCTGATGCCT TTGCACCTAA CGCCAAAATT GTACATGTAG ATATTGATCC TTCAGAAATC	6420
	AATAAAGTTA TTCATGTAGA TTTAGGTATT ATTGCAGACT GTAAAAGATT TTTAGAATGT	6480
	TTAAATGATA AAAATGTTGA GACTATAGAA CACAGTGAAG GGGTTAAACA TTGTCAAAAT	6540
10	AATAAGCAGA AACACCCATT TAACTTGGT GAAGAAGATC AAGTATTTTG TAAGCCACAA	6600
	CAACAATCG AATATATCGG CAAAATTACA AATGGTGAAG CAATTGTTAC TACAGACGTG	6660
15	GGACAACATC AAATGTGGG AGCTCAATTT TATCCATTTA AAAATCACGG ACAATGGGTT	6720
	ACAAGCGGTG GTTTAGGAAC AATGGGATTC GGTATTCCTT CGTCAATTGG TGCCAAATTA	6780
	GCTAATCCTG ATAAAACAGT CGTATGTTTC GTCGGTGACG GTGGTTTCCA AATGACAAAC	6840
20	CAAGAAATGG CACTTTTACC CGAATATGGT TTAGATGTCA AAATCGTACT AATCAATAAT	6900
	GGAACATTAG GTATGGTTAA ACAATGGCAA GATAAGTTCT TTAATCAACG CTTCTCACAC	6960
	TCAGTATTTA ATGGTCAACC TGATTTTATG AAAATGGCAG AAGCATATGG CGTCAAAGGT	7020
25	TTCTTAATCG ATAAGCCAGA ACAACTGGAA GAACAATTAG ATGCAGCGTT TGCTTATCAA	7080
	GGACCAGCTT TAATTGAGGT TCGTATTTCC CTTACTGAAG CTGTAACCCC AATGGTTCGG	7140
	AGTGGCAAAT CAAATCATGA AATGGAGGGC TTATAATGAC AAGAATTCTT AAATTACAAG	7200
30	TTGCGGATCA AGTCAGCACG CTAAATCGAA TTACAAGTGC TTTTGTTGCG CTACAATATA	7260
	ATATCGATAC ATTACATGT ACACATTCTG AACAACTGG GATTCTAAC ATGGAAATTC	7320
35	AAGTCGATAT TCAAGATGAT ACATCACTTC ATATATTAAT TAAAAATTA AAACAACAAA	7380
	TTAATGTTTT AACGGTTGAA TGCTACGACC TTGTTGATAA CGAAGCTTAA TTTTAAGACA	7440
	AAGGCAATGA TGCGCTAATT AGTTATAGAT ATATCATAGG CTGCTAGTTA ACATCTGCCA	7500
40	CTATTACAAA GTTATATTTT AGAATTTTCG AAACACAAAA TATTTAATTA TTTGGAGGAA	7560
	TTTATTATGA CAACAGTTTA TTATGATCAA GATGTAAAA CGGACGCTTT ACAAGGCAAA	7620
	AAAATTGCAG TAGTAGGTTA TGGATCACAA GGTACGCGC ATGCACAAAA CTAAAAGAC	7680
45	AATGGATATG ATGTAGTCAT CGGCATTGCG CCAGGTCGTT CTTTTGACAA AGCTAAAGAA	7740
	GATGGATTTG ATGTGTTCCC TGTTCAGAA GCAGTTAAGC AAGCTGATGT AATTATGGTG	7800
	CTATTACCTG ATGAAATTCA AGGTGATGTA TACAAAAACG AAATTGAACC AAATTTAGAA	7860
50	AAACATAATG CGCTTGCAAT TGCTCATGGC TTAAACATTC ATTTTGGTGT TATTCAACCA	7920
	CCAGCTGATG TTGATGTATT TTTAGTAGCT CCTAAAGGAC CGGGTCATTT AGTTAGACGT	7980
55		

	CAAGCACGTA ATATTGCTTT AAGTTATGCA AAAGGTATTG GTGCAaCTCG TGCAGGTGTT	8100
	ATTGAAACAA CATTTAAAGA AGAACTGAG ACAGATTTAT TTGGTGAACA AGCAGTACTT	8160
5	TGCGGTGGTG TATCGAAATT AATTCAAAGT GGCTTTGAAA CATTAGTAGA AGCGGGTTAT	8220
	CAACCAGAAT TAGCTTATTT TGAAGTATTA CATGAAATGA AATTAATCGT TGATTTGATG	8280
	TATGAAGGCG GTATGGAAAA TGTACGTTAC TCAATTTCAA ATACTGCTGA ATTTGGTGAC	8340
10	TATGTTTCAG GACCACGTGT TATCACACCA GATGTTAAAG AAAATATGAA AGCTGTATTA	8400
	ACTGATATCC AAAATGGTAA CTTCAGTAAT CGCTTTATCG AAGACAATAA AAATGGATTC	8460
	AAAGAATTTT ATAAATTACG CGAAGAACAA CATGGTCATC AAATTGAAAA AGTTGGTCGT	8520
15	GAATTACGCG AAATGATGCC TTTTATTAAA TCTAAAAGCA TTGAAAAATA AGATAGACCT	8580
	ACAATGAGGA GTTGTAAAT ATGAGTAGTC ATATTCAAAT TTTTGATACG AACTAAGAG	8640
20	ACGGTGaACA AACACCAGGA GTGAATTTTA CTTTGATGA ACGCTTGCGT ATTGCATTGC	8700
	AATTAGAAAA ATGGGGTGTA GATGTTATTG AAGCTGGATT TCCTGCTTCA AGTACAGGTA	8760
	GCTTTAAATC TGTCAAGCA ATTGCACAAA CATTAAACAAC AACGGCTGTA TGTGGTTTAG	8820
25	CTAGATGTAA AAAATCTGAC ATCGATGCTG TATATGAAGC AACAAAAGAT GCAGCGAAgC	8880
	CGGTcGTGCA TGTTTTTATA GCAACATCAC CTATTCATCT TGAACATAAA CTTAAATGT	8940
	CTCAAGAAGA CGTTTTAGCA TCTATTAAAG AACATGTCAC ATACGCGAAA CAATTATTTG	9000
30	ACGTTGTTCA ATTTTCACCT GAAGATGCAA CGCGTACTGA ATTACCATT CTTAGTGAAAT	9060
	GTGTACAAAC TGCCGTTGAC GCTGGAGCTA CAGTTATTAA TATTCCTGAT ACAGTCGGCT	9120
	ACAGTTACCA TGATGAATAT GCACATATTT TCAAAACCTT AACAGAATCT GTAACATCTT	9180
35	CAAATGAAAT TATTTATAGT GCTCATTGCC ATGACGATTT AGGAATGGCT GTTTCAAATA	9240
	GTTTgCTGC AATTGAAGGC GGTGCGAGAC GAATTGAAGG CACTGTAAAT GGTATTGGTG	9300
40	AACGAGCAGG TAATGCAGCA CTTGAAGAAG TCGCGCTTGC ACTATACGTT CGAAATGATC	9360
	ATTATGGTGC TCAAACCTGCT CTTAATCTCG AAGAACTAA AAAACATCG GATTTAATTT	9420
	CAAGATATGC AGGTATTCGA GTGCCTAGAA ATAAAGCAAT TGTTGGCCAA AATGCATTTA	9480
45	GTCATGAATC AGGTATTCAC CAAGATGGCG TATTAAACA TCGTGAAACA TATGAAATTA	9540
	TGACACCTCA ACTTGTTGGT GTAAGCACGA CTGAACTCC ATTAGGAAAA TTATCTGGTA	9600
	AACACGCCTT CTCAGAGAAG TTAAAAGCAT TAGGTTATGA CATTGATAAA GAAGCGCAAA	9660
50	TAGATTTATT TAAACAATTC AAGGCCATTG CGGACAAAAA GAAATCTGTT TCAGATAGAG	9720
	ATATTCATGC GATTATTCAA GGTCTGAGC ATGAGCATCA AGCACTTTAT AAATTGGAAA	9780
55		

	AAGAGGGTCA	TATTTACCAG	GATTCAAGTA	TTGGTACTGG	TTCAATCGTA	GCAATTTACA	9900
	ATGCAGTTGA	TCGTATTTTC	CAGAAAGAAA	CAGAATTAAT	TGATTATCGT	ATTAATTCTG	9960
5	TCAGTGAAGG	TACTGATGCC	CAAGCAGAAG	TACATGTAAA	TTTATTGATT	GAAGGTAAGA	10020
	CTGTCAATGG	CTTTGGTATT	GATCATGATA	TTTTACAAGC	CTCTTGTAAG	GCATACGTAG	10080
	AAGCACATGC	TAAATTTGCA	GCTGAAAATG	TTGAGAAGGT	AGGTAATTAA	TTATGACTTA	10140
10	TAACATTGTT	GCCCTACCTG	GTGATGGAAT	CGGTCCAGAA	ATTTTGAACG	GATCTCTATC	10200
	ATTGCTTGAA	ATTATAAGTA	ATAAATATAA	CTTTAATTAT	CAAATAGAGC	ACCACGAATT	10260
	TGGTGGTGCC	TCTATTGATA	CATTCGGCGA	GCCTTTAACT	GAGAAAACCT	TAAATGCGTG	10320
15	TAAAAGAGCA	GATGCTATTT	TACTGGGTGC	AATCGGTGGA	CCTAAATGGA	CAGATCCTAA	10380
	CAATCGACCA	GAACAAGGAT	TATTAATAAT	GCGTAAATCC	TTAAATTTAT	TTGTAAATAT	10440
20	ACGCCCCACT	ACCGTTGTCA	AAGGCGCTAG	TTCTTTATCA	CCTTTAAAGG	AAGAACGCGT	10500
	TGAAGGCACA	GATTTAGTTA	TAGTCCGTGA	ATTGACAAGT	GGTATTTATT	TTGGAGAACC	10560
	TAGACATTTT	AATAATCACG	AGGCCTTAGA	TTCTCTTACT	TATACAAGAG	AAGAAATAGA	10620
25	ACGCATTGTT	CACGTAGCAT	TTAAATGGC	CGCTTCAAGA	CGAGGAAAAC	TAACATCAGT	10680
	TGATAAAGAA	AATGTATTAG	CTTCTAGTAA	ATTGTGGCGC	AAAGTCGTAA	ATGAAGTAAG	10740
	TCAATTATAT	CCAGAAGTAA	CAGTAAATCA	CTTATTTGTT	GATGCTTGTA	GTATGCATTT	10800
30	AATCACAAAT	CCAAAACAAT	TTGACGTCAT	CGTATGTGAA	AACTTATTTG	GCGATATTTT	10860
	AAGTGATGAA	GCTTCAGTGA	TTCCTGGTTC	ACTTGGTTTA	TCACCTTCTG	CTAGTTTTAG	10920
	TAACGATGGT	CCAAGATTGT	ATGAGCCTAT	TCATGGATCA	GCACCAGATA	TTGCAGGTAA	10980
35	AAACGTTGCC	AATCCATTTG	GAATGATTCT	ATCTTTAGCG	ATGTGTTTAC	GTGAAAGCTT	11040
	AAATCAACCA	GATGCTGCAG	ATGAATTAGA	ACAACATATT	TATAGCATGA	TTGAACATGG	11100
40	GCAAACGACA	GCAGATTTAG	GCGGCAAATT	GAATACTACT	GATATTTTCG	AAATTCTATC	11160
	TCAAAAATTG	AATCACTAAG	GGGGAGATGT	AAATGGGTCA	AACATTATTT	GACAAGGTGT	11220
	GGAACAGACA	TGTGTTATAC	GGGAAATTGG	GCGAACCGCA	ACTATTATAC	ATTGATTTAC	11280
45	ACCTTATACA	TGAAGTTACT	TCTCCTCAAG	CATTTGAAGG	ACTTAGGCTT	CAAAACAGAA	11340
	AATTAAGACG	CCCAGATTTA	ACATTTGCAA	CACTCGATCA	CAATGTTTCT	ACTATTGATA	11400
	TATTCAATAT	TAAAGATGAA	ATTGCAAAAC	AACAAATCAC	AACATTACAA	AAAAACGCCA	11460
50	TAGATTTTGG	GGTGATATTT	TTTGATATGG	GTTCTGATGA	ACAAGGTATT	GTTTACATGG	11520
	TAGGACCTGA	GACAGGACTT	ACACAGCCTG	GCAAGACAAT	CGTTTGTGGT	GACTCTCACA	11580
55							

## EP 0 786 519 A2

	ATGTTTTTCGC	AACTCAAACG	CTATGGCAAA	CAAAACCCAA	AAACTTAAAA	ATCGATATTA	11700
	ATGGTACCTT	ACCAACAGGC	GTCTATGCTA	AGGACATTAT	TCTGCATTTA	ATTAAAACGT	11760
5	ATGGTGTGGA	CTTTGGTACA	GGCTATGCTT	TGGAATTTAC	TGGCGAAACA	ATTAAAAACC	11820
	TTTCAATGGA	TGGTCGAATG	ACTATTTGTA	ACATGGCTAT	CGAAGGTGGT	GCCAAATACG	11880
	GCATAATCCA	ACCTGATGAT	ATAACATTG	AATATGTTAA	AGGGAGACCA	TTTGCCGATA	11940
10	ACTTCGCTAA	ATCAGTTGAT	AAGTGGCGTG	AgCTATATTC	TGATGACGAC	GCGATATTTG	12000
	ATCGTGTAAT	TGAACCTGAT	GTTTCAACAT	TAGAACCACA	AGTGACATGG	GGAACATAATC	12060
	CTGAAATGGG	TGTTAATTTT	AGTGAACCAT	TCCCTGAAAT	CAATGATATC	AACGATCAAC	12120
15	GTGCGTATGA	TTATATGGGG	TTAGAACCAG	GTCAAAAAGC	TGAAGACATC	GACTTAGGGT	12180
	ATGTTTTTCT	CGGTTTATGT	ACAAATGCTA	GACTATCAGA	TTTGATTGAA	GCTAGTCATA	12240
20	TTGTAAAGG	AAATAAAGTT	CATCCAAATA	TTACAGCTAT	TGTCGTACCA	GGTTCTCGTA	12300
	CAGTAAAAAA	AGAAGCAGAA	AAATTAGGTC	TAGATACTAT	CTTTAAAAAT	GCAGGATTTG	12360
	AATGGCGTGA	ACCAGGATGT	TCAATGTGTT	TAGGCATGAA	TCCTGACCAA	GTACCTGAGG	12420
25	GCGTACATTG	TGCATCTACA	AGTAATCGAA	ACTTTGAAGG	ACGACAAGGC	AAAGGTGCAA	12480
	GAACACATTT	AGTATCCCCt	GCTATGGCAG	CAGCAGCAGC	TATTCATGGT	AAATTTGTGG	12540
	ACGTAAGAAA	GGTGGTTGTT	TAAATGGCAG	CAATCAAACC	TATTACAACA	TATAAAGGTA	12600
30	AAATAGTCCC	TCTCTTCAAC	GACAATATCG	ATACAGACCA	AATCATTCCT	AAGGTACACT	12660
	TAAAGCGTAT	TTCAAAAAGT	GGCTTTGGTC	CATTTGCTTT	TGATGAATGG	CGGTACTTAC	12720
	CTGATGGTTC	AGATAATCCT	GATTTCAATC	CTAACAAACC	ACAATATAAA	GGGGCTTCTA	12780
35	TTTTAATTAC	TGGAGATAAT	TTTGATGTG	GTTCAAGTCG	TGAACATGCT	GCTTGGGCTC	12840
	TTAAGGACTA	TGGTTTTTCAT	ATTATTATTG	CAGGAAGTTT	CAGTGACATA	TTTTATATGA	12900
40	ATTGCACTAA	AAATGCGATG	TTGCCTATCG	TTTTAGAAAA	AAGTGCCCGT	GAACATCTTG	12960
	CACAATATGT	TGAAATTGAG	GTCGATTTAC	CAAATCAAAC	TGTGTCATCA	CCAGACAAGC	13020
	GTTTCCATTT	TGAAATTGAT	GAAACTTGGA	AGAATAAACT	TGTAAATGGC	TTAGATGACA	13080
45	TTGCAATCAC	CCTACAATAT	GAATCATTA	TAGAAAAATA	TGAAAAATCa	CTTTAAGGGA	13140
	GTTGAATATT	ATGACAGTCA	AAACAACAGT	TTCTACGAAA	GATATCGATG	AGGCATTTTT	13200
	AAGACTTAAA	GATATTGTCA	AAGAAACACC	TTTACAATTA	GACCATTACT	TATCTCAAAA	13260
50	GTATGATTGT	AAAGTCTATT	TAAAACGAGA	AGATTTACAA	TGGGTACGTT	CTTTTAAATT	13320
	AAGAGGTGCT	TACAACGCTA	TTTCTGTTTT	ATCAGATGAA	GCTAAAAGTA	AAGGTATTAC	13380

55

EP 0 786 519 A2

	AAACGCTGTT ATCTTTATGC CAGTCACTAC ACCTTTACAA AAGGTAAATC AAGTAAAGTT	13500
	CTTTGGAAAT AGTAACGTTG AAGTTGTA CTCTGGTGAT ACATTGATC ACTGTTTAGC	13560
5	TGAAGCTTTA ACTTATACAA GTGAACATCA AATGAACTTT ATAGATCCAT TCAATAATGT	13620
	TCATACAATT TCTGGACAAG GTACGCTTGC TAAAGAAATG CTAGAACAAG CAAAGTCTGA	13680
	CAATGTTAAC TTTGATTATC TATTTGCCGC AATTGGTGGT GGCGGTTTAA TTTCAGGTAT	13740
10	TAGTACTTAC TTTAAACCT ATTCACCTAC CACGAAAATT ATAGGTGTTG AACCTTCAGG	13800
	TGCAAGTAGT ATGTATGAAT CTGTTGTGGT AAATAATCAG GTAGTCACAT TGCCTAATAT	13860
	CGATAAATTT GTGGACGGTG CATCTGTAGC TAGAGTTGGC GATATTACAT TTGAAATTGC	13920
15	AAAAGAAAAT GTAGATGATT ACGTTCAGT AGATGAAGGT GCAGTTTGTT CTACGATTTT	13980
	AGATATGTAT TCAAAACAAG CAATTGTAGC AGAACCTGCT GGCGCATTAA GTGTAAGTGC	14040
20	GCTTGAAAAC TATAAAGATC ATATTAAAGG TAAACAGTG GTTTGTGTCA TTAGTGGTGG	14100
	TAATAATGAT ATTAATCGAA TGAAAGAAAT TGAAGAACGT TCATTACTAT ACGAAGAAAT	14160
	GAAGCATTAC TTTATCTTAA ATTTCCCTCA ACGTCCAGGT GCATTGAGAG AATTGTAAA	14220
25	TGACGTATTA GGACCTCAAG ACGATATTAC TAAATTGAA TACTTAAAAA AATCTTCTCA	14280
	AAATACAGGT ACTGTCATTA TTGGTATTCA ACTTAAAGAT CATGATGATT TAATACAAC	14340
	CAAACAACGT GTAAATCATT TCGATCCTTC CAATATTTAT ATTAATGAAA ATAAGATGTT	14400
30	ATATTCATTG TTAATTTAAC ACATAGTAAG AAAACAGTC ATAAATTGAT TTCTAATTGA	14460
	AATCATCTTA TGA CTGCTTTT TTATTATACT TTACATTTCT CGTTTCGTCA GATTCAAACG	14520
	TTTTCACTTC GCCAAGCCAT CTTTCTTTGT GTTTGCTTTT aTTTGTACGT TTTAGACATA	14580
35	AAAAAaGAGA CCTTGCGGTC TCAATGCGGC TCATCGCATC CACTTTTTCG CTGGCAACGT	14640
	TCTACTCTAG CGGAACGTAA GTTCGaCTAC CATCGACGCT AAGGAGCTTA ACTTCTGTGT	14700
40	TCGGCATGGG AACAGGTGTG ACCTCCTTGC TATAGTCACC AGACATATGA ATGTAATTTA	14760
	TACATTCAAA ACTAGATAGT AAGTAAAAGT GATTTTGCTT CGCAAAACAT TTATTTTGAT	14820
	TAAGTCTTCG ATCGATTAGT ATTCGTCAGC TCCACATGTC ACCATGCTTC CACCTCGAAC	14880
45	CTATTAACCT CATCATCTTT GAGGGATCTT ATAACCGAAG TTGGGAAATC TCATCTTGAG	14940
	GGGGGCTTCA TGCTTAGATG CTTTCAGCAC TTATCCCGTC CACACATAGC TACCCAGCTA	15000
	TGCCGTTGGC ACGACAACGT GTACACCAGA GGTATGTCCA TCCCGGTCCT CTCGTACTAA	15060
50	GGACAGCTCC TCTCAAATTT CCTACGCCCA CGACGGATAG GGACCGAACT GTCTCACGAC	15120
	GTTCTGAACC CAGCTCGCGT ACCGCTTTa TGGGCGAACA GCCCAACCCT TGGGACCGAC	15180
55		

GTGGAACTT

15249

## (2) INFORMATION FOR SEQ ID NO: 103:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14051 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

15	GTGGCAATAT TTCTAGTTCT CGTTTTGATA AGATTTTAAA AGGATCTGTT GTGTTTGCAG	60
	TGTCCTGATT TGAATTAGAT ACAAATTCAT TCACTAAAGA TGTTGTAAGT TTCATATCTA	120
	CATATGTTTC ACCTTTATAT ACAGTTCGAA TAGCTAACAA TAATTGTTCA TCAGGTGCAT	180
20	TTTTCAATAT GTAACCTTTC GCACCATTAC GCAACACATG GAACAAATAC TCCTCATCAT	240
	CAAACATTGT TAATATTAGT ATTTTAGTTT CAGGAAAAC TGCAGCAATT TTA CTCTAGT	300
	CGATAAGACC TGA CTACCT GGTGGC ATAC TTAAATCCAT TAGTAACACA TCAGGTTTAT	360
25	ATTCCATTAC TTTTGGTAA GCTTCGACGC CATCTGCAGC CGTTGCAACA ACTTCCATAT	420
	CATTTTGATA ATTTAAATC ATAGAGAACC CCGTACGGAC AACAGCGTGA TCATCGGCAA	480
	TGACTATTTT CAATTTTATT CCCCCAATGT ATGTTTCAAA TTGGAATGTT CAATGTAACA	540
30	TTGGTACCCT CACCAATTTT CGTTTCAATA TTGACGCTAC CGCTGACTAA CTCAGCTCGC	600
	TCATTCATTC CATATAAACC GAGTCCAGAA CCTTTAGGCT TAGAACTTGG ATCAAAAACCA	660
	TTTCCCGCAT CTATCACTTC TGCTACCAA TGGCGCCCAG TTTGACGGAT ACCTACATTT	720
35	ATTTCAATTA CATCAGCGTA TTTCAACGCA TTTAAAATAG CTTCTTGCAC TACTCGATAA	780
	ACAACCGTTT CAATATCACT ATCAAAGCGA GTATTTTAA TATTTGATGT ATATATGATT	840
40	TTTATTCAT AATTTTCTTC AAAGTGT TTA AAATATGATT TAAAAGCTGC TTCAAGGCCT	900
	AGATCATCCA AAGAAGCGGG TCTTAATTCA ACCGACATAT TACGTATATC ATCAATTAAT	960
	TTAGCGACAA TATATTCAAT ATTTTCTGCG TCTTCCAAA GCTTAGTTGT ATCTTCTTGA	1020
45	TATTTTAATA ATCTCAATTG AACATCTACA TTGAGCATTT CTTGAATCAC ACTATCATGT	1080
	AACTCTCTAG AAATTCGCTT TCTTTCATTT TCTTGGGCTG AGATTGTTTT ACGCATCATA	1140
	CGTTGTTGAT GCAATTTCTC TTGCTGTTCA ATTTGTGATG AAACATTTTG AAGCGTAAAT	1200
50	GCAATGAATC CCCTGTCTTG ATCAATCAAC TGATATGTTG CTGTAAATGG CATCACTTTT	1260
	TGATCTTTTG TCTTCATAAA TACTTGGAAA TTCGTAGCTT GTACTTGCAT CGATTCTAAG	1320



EP 0 786 519 A2

	ATCGCATTCTG CCACAGCACT GTAATTATCT TCTTCAGATA ATATATCTTT AGCAGCATCA	1440
	TTTATTGCAA TAATTTTACC GTTATCATCA GCAAAACTA TCTTTTCGAT TGAATGCTCA	1500
5	TAATATTTTT TCAATAAAGT ATCTAACTGT ATACTGTCCT CATTAAATCAT GACTTACACC	1560
	CTAATTCATC TCATTATTTA TCATCATTGA AAATACCAA CTTACGTTGA ATATCATCAT	1620
	TATCAAATAT TTTTGGTAAA GGACGACCAT CTCTTTGACC AAATAATAGT ACGCCATACA	1680
10	CTTGATTCTT ATACCAAAGC GGCAGTCTA AACTGCTGT TAATGATTCTG CTCAATAAAA	1740
	TTGGATAGTC AATCTTTTCT TCAGGCCCTA AAGCTAAACC AACATTGGCT ATTACCATAC	1800
	GCTTTCCTGT TTTTATAACA GTTCCAGCTA ATCCACGACC TTTTCTTAAA ATAATCAATT	1860
15	TAAATCGATT ATTTTATTA CCTGAAACAT AGTGCCATTT TATTGGAGAT GATGGTTTGT	1920
	TAGATTCATA GAAAGCGATT GCCGCAAAT CATAACCCTC TTCTTTGCGT ATTTTATCTA	1980
20	ATGTCTCTTG AAATCTACGA TCTTCAATTA TTGCTTCTGG TGTCAAATCC TTTCACCTCT	2040
	TATGCTTACA CTTTATTCTT ACGGTAAATA ATATATCTGC GATTTATATA TGTCAAAGGT	2100
	ACACTCCAAA CATGCACCAA ACGTGTAAT GGCCAACAAG CCATAATAGT GAAACCTAAC	2160
25	AATATATGCA TTTTAAATGC AATCGGCACA CCACTCATCA ATGACGCATC TGGTTTTAAC	2220
	ATAAATAATT GTCTAAACCA AATTGATAAT GAAGTTCTGT AGTTAAAGTC TGGATGTTGT	2280
	ATATTTGTTA CTAATGTTGC GTAACATCCC ATAAATACGA TAAGTAATAA TAAGAAATTT	2340
30	ACAAATATAT CCGACGCTGA ACTTAATCTT CGAATACTTT TCGTAGTAAC ACGTCTCGCT	2400
	GTTAATAAAA ACATCCCTAT CAAAGTTATT ATACCAAAGA TGCTACCAAT ATAAACAGCG	2460
	CCTATATGAT ATAAATGCTC AGACACACCC ACTGCATCCA TCCATGGTTT CGGTATTAAC	2520
35	AATCCAACCTA CGTGTCAAA AAACACTGGA ATAATACCTA AGTGAAATAA TAAACTTCCC	2580
	CACATCAACC TTTTCTTTC TATTAATTCA CTAGATTTAG CTGTCCAAGA AAATTTATCA	2640
	TAACGATAAC GTGCAATATG ACCTGCGACA AAGACAACTA AACATAAATA CGGAAATATA	2700
40	ACCCATAAAA ACTGATTAAG CATGATGTTT CACTCCTTTT GGTGATGTCA AACATAATTT	2760
	CAATGTTTTT CTAAGTGCTT GAATCACATA GGCATATGGA TTGTTATCTT CACCAAGTGC	2820
45	ATTGCGCATC ACATATGTTT CATCCTCAAT AATCATAATG ATTAATTGAA TATTCTCTTC	2880
	AGCTCTTGGA TCATTCGCC ATTCTGCCAC TTGCAAAAAT TGAAGCATCA ACGGTAGATA	2940
	ATCAGAAAGT TCATTATCTA CCATTTCTAG TCCAAACATT TCATATAATA CCTTTAATTT	3000
50	AGCTAACATT TGCCACGTT CTTTTGCGT ATCAAATTTG TTATACGTCA TATATAATGG	3060
	TGCTTTTTTC GTAAATCAA ATGTATCTGT ATAAATCGCT TTGATTTCTG ATAATGAAAA	3120
55		

	TGTTTCTTCA	AAAGTTTTTG	GATGAAAAGT	TAATTTTTCT	GGAAAACATA	ACTGTTGTGC	3240
	CATATATCCA	AAACTTTCTT	GATATTTTTT	AAAATTATCG	AAATTAATCA	CGGAAAATCC	3300
5	CTCCATAGAA	ATTCTCATTA	TAAATTTCTT	GACCAGTTTT	CCCTGAACCT	ACTGCAACGC	3360
	CACAGCCTTC	ACAGTTATCT	CCAAAATGCT	CGCCGCCGTA	ATTGTATCCT	GTACTIONCTT	3420
	GTGCGTGATA	CGTATCTAAA	TAGGTTTCTT	TGTGTGATGT	TGGAATAACA	AATCGATCTT	3480
10	CATATTTGGC	TAGTCCTAAT	AAACGATACA	TGTCTTTAGT	TTGGCGCTCG	GTTATACCTA	3540
	ATCGCTCTAA	TCGAGACGTG	TCAAATGGCT	GTTGAGTAAC	TTGAGATCTC	ATATAACTTC	3600
	TCATCATTGC	CATACGTTGT	AGGGCTCCTT	TTACTGGCTC	TGTATCTCCT	GCAGTGAAAA	3660
15	TATTAGCTAA	GTATTCAATA	GGTAAACGCA	TTTCTTCAAT	GGCTGGGAAA	ATCGCATCTG	3720
	GATTTTGAGT	TGTATTTTTA	CCTTCAAAAT	AGCTCATAAT	TGGGCTAAGT	GGTGGGCAAT	3780
20	ACCAAACCAT	CGGCATCGTT	CTAAATTCAG	GATGTAACGG	AAATGCAAGT	TTATATTCAA	3840
	TTGCTAACTT	ATAAATIGGA	GAGTTTTGTG	CAGCTTCAAT	CCAATCGTAA	CCAATACCAT	3900
	CTTTTTCAGC	TTGAGCAATG	ACTTCTTCGT	CAAATGGGTT	TAAGAATATA	TCTAATTGTT	3960
25	TTTCATATAA	ATCTTTCTCG	TCTACTGCTG	AAGCTGCTTC	ATGAACTCGA	TCTGCATCAT	4020
	ATAATAAAAC	ACCTAAGTAA	CGCATACGTC	CTGTACAAGT	TTCAGAGCAT	ACCGTAGGCA	4080
	TACCCGCCTC	GATTCTCGGG	AAACAGAAAG	TACACTTTTC	AGCTTTGTTC	GTTTTCCAAT	4140
30	TGAAGTAAAC	TTTCTTATAT	GGACAACCTG	TCATACAGTA	ACGCCATCCA	CGACATGCGT	4200
	CTTGGTCAAC	TAATACAATG	CCATCTTCAT	CACGTTTATA	CATAGCACCT	GAAGGACACG	4260
	ATGCAACGCA	ACTTGGATTG	AAGCAATGTT	CACATAAACG	TGGTAAATAC	ATCATAAAAG	4320
35	TTTCGTCAAA	TTGGAATTTA	ATATCTTCTT	CTATTTTTTG	GATGTTAGGA	TCTTTTGGAC	4380
	CTGTAACATG	ACCACCTGCT	AAGTCATCTT	CCCAGTTAGG	TCCCCATTCA	ATTTCAATGT	4440
40	TATCCCCCGT	AATTTCTGAA	TACGCTCTAG	CAACTGGCGA	ATGCTTCCCT	GATTTTCGCAG	4500
	TTGTTAAATG	TTCATAATTA	TAGTTCCATG	GCTCATAATA	ATCTTTAATT	AATGGCATAT	4560
	CTGGGTATATA	AAAAATTTTA	CCTAAAGCAA	TTTTTGAAAT	TCTACTTCCA	GATTTTAATT	4620
45	CAAGTTTCCC	TTTACGATTT	AGTACCCAAC	CACCTTTGTA	GTGTTCTTGG	TCTTCCCAAC	4680
	GTTTCGGATA	CCCTACACCT	GGCTTCGTTT	CTACGTTGTT	GAACCACATG	TACTCAGCAC	4740
	CTGGACGATT	TGTCCaAGTG	TTTTTACATG	TCACACTACA	CGTATGGCAT	CCTATGCATT	4800
50	TATCTAAATT	TAATACCATC	GCAACTTGCG	CTTTAATCTT	CAAGCCAATT	AACCTCCTTC	4860
	ATCTTTCTAA	CTGCTACATA	TAAATCCCTT	TGGTTCCCAA	TTGGTCCATA	ATAATTAAAG	4920
55							

	GGCGCGTTGT GTGAACCACC ACGTGTATCT GTAATTTCTG ACCCAGGCGT TTGAATATGT	5040
	TTATCTTG TG CATGATACAT AAACATTGTA CCTTTAGGCA TACGATGCGA AATAACTGCT	5100
5	CTTGCCGTTA CAACACCATT ACGGTTATAC ACTTCTAGCC AATCATTATC TTGGATATCG	5160
	TGTTTTTTCAG CATCTTCATT TGATATCCAA ACCGTTGGAC CACCTCTAAA TAGTGTCAAC	5220
	ATATGCTTAT TATCTTGATA CATTGAGTGT ATATTCCATT TTCCATGAGG CGTTAAATAA	5280
10	CGCAgTACCA AAGCATCTGT ACCACCTTTA ATTTTCTTAT CTCTATTCCC AAATACCATT	5340
	GGCGGCAATG TCGGTTTATA TACTGGTAAG CTCTCCCCAA ATTGTTGGAA AACTTCGTGA	5400
	TCCACATAAT AACTTTGACG TCCTGTTAAT GTTCTAAAAG GTACTAGACG TTCTATATTC	5460
15	GTTGTAAATG GTGAATATCG TCGACCTTGT TTATTTGAAC CTGGGAATAC TGCTGTCGGT	5520
	ATTACTTCTC GTGGTTGTGA AGTTATATTT AAAAACGAAA TTTTCTCAGC AGCGCGTTCTG	5580
20	CTAGAAATAT CTTTTAACGG CATTCCAGTT TGTTCTTCGA GATCTTCATA TGATTTTTGT	5640
	GATAATTTAC CATTTCGTAGC AGATGAAATA CTTAGTATTG CATCAGCTAC ATTACGTGCT	5700
	GTATCAATAC GTGGACGATT CGCTCTCACA GAATCATCAT TTGTATCACT CCACGTACCT	5760
25	AACATACTTT TTAATTCTTC ATATTGTTCA CTGACACCGA AACTTACACC ATGTGCTCCA	5820
	ACTTTCCCTT TTTCAAGTAC AGGACCAAGC GTGACATATT TGTCGTAAAT TTTAGTGTAG	5880
	TCGCGTTCTA CAATTGCAAA GTTAGGCATT GTACGTCCAG GTACCGCTTC AATTTCACCC	5940
30	TTTCGACCAAT CTTTCACTAC GCCGTATGGT GTTGAAATTT CTGCTTTGT ATCATGACTA	6000
	AGTGGAGTTG TCACAACATC TTTAAACGTT CCAGGTAAAT AGTCTTTTGC CATTTCTGAA	6060
	AATGCTTTTG CCAACGTTTT ATAAATATCC CAGTCTGAAC GCGATTCCCA TAACGGATCA	6120
35	ATGGCAGGAT TGAAAGGATG TACATATGGA TGCATATCCG TTGATGATAA ATCATGTTTT	6180
	TCATACCAAG TCGCTGCCGG CAAAACAATG TCAGAATATA ACGGTGTTGC CGTCATTCTG	6240
40	AAGTCTAAAG AGACCACTAA ATCTAACTTA CCTGTTGTTT CTTCACGCCA CGTAATTTCT	6300
	TCTGGCTTTT CATCTTCATT TGGTGTAGCT AATAACCTG ATTTTGTGCC AAGTAAATGC	6360
	TTCATAAAGT ATTCTTGACC TTTTGCAGAA CTTGAAATTA AGTTTGAACG CCATATAAAT	6420
45	AATGATTTTG GATGATTCTT TTTCAAATCA GGATCTTCTA TTGCAAATTG TGTTTGTTTT	6480
	GATTTCACTT CATCAATTGC ACGTTGCAAA ATCGCTTCAT TTGAATCTAT ACCTTCATCT	6540
	TTAGCTTCTT CTGCAACAA CAACTATTT TTATTAAATT GTGGATATGA TGGTAACCAA	6600
50	CCAAGTCTAG CTGCTAAAAC ATTATAATCA GCTGGATGTT GATGCTTTAA CTCCTCTGTT	6660
	TTAGCTAATG GAGATTTTAA ACGATCTACA TTTGACTCTT CATATTTCCA TTGGTCTGTT	6720

55

	AATGCGACAG TACTCCATCC TTCAATCGGA CGACATTTTT CTTGTCCCAC ATAGTGAGCC	6840
	CAACCGCCAC CATTACACCC TTGACAGCCA CATAACATAA CTAAGTTTAA GATTGAACGA	6900
5	TAAATCGTAT CTGAGTTAAA CCAATGGTTA ATACCCGCAC CCATGATAAT CATTGAACGC	6960
	CCTTCAGTAT CGATAGCGTT TTGCGCAAAT TCTTTCGCTA CTTGAATGAC AACACTTTGT	7020
	TTTACGCCTG AAATGGCTTC TTGCCAAGCA GGTGTATATT TTGATTCTGC ATCGTCGTAT	7080
10	CCTTTTGATT CTAATTTATG ATCAAAACGA CGCACGCCAT ATTGACTTGC CATTAAGTCA	7140
	AAAATTGTAG CAATACGGAC TTTGTCACCA TTTGCTAAAG TGACTTGTGC AGTTGGAATT	7200
15	GGACGATTGA ATATCCCATC TCCATCACTA TCAAAGTATG GGAATTGAAT TGTTCCTAAT	7260
	TCGTATCCAC CTTCTGTCTAT TGATAATGTA GGGTTAATTT TAGAACCATC TTCTGTTTCT	7320
	AGTTTTAAGT TCCACTTCTT ACCTTCTTCC CAACGTTGAC CCATTGTGCC ATTAGGTAAT	7380
20	ACTAAACTAT CGCTGATTGC ATCATGAATA ACTGGCTTCC ATTCGCCTTG CTCTGTTGTT	7440
	TGACCTAAGT CACTCGCTCT TAAAAATCGA CCCGCTTTAT ATCCATTTTC ATCTTCATCC	7500
	AGCATGATAA GAAACGGCAT ATCTGTATAT TGTTTAGCGT AATTTATAAA GCGTTCATTA	7560
25	GGTTGATTAA CATAATGTTC TTGTAAAATA ACATGCGTCA TTGCTTGTGC AATTGCAGCA	7620
	TCTGAACCAG GATTCCGTGC TAGCCAGTTA TCTGCAAATT TCACATTTTC TCGGTAATCT	7680
	GGTGCTACTG AAATGACTTT TGTACCTTTA TAGCGGACTT CAGTCATAAA ATGTGCATCC	7740
30	GGAGTACGTG TTAAAGGTAC ATTAGAGCCC CACATAATAA TGTATGATGC GTTATACCAG	7800
	TCACTTGATT CAGGCACATC TGTTCGTCTT CCCCATAATT GTGGAGAGGC AGGTGGTAAA	7860
35	TCTGCATACC AGTCATAAAA ACTAAGCATT TCACCACCAA GCAAATTGAT GAATCGAGCA	7920
	CCTGCTGCAT AACTAATCAT TGACATCGCT GGAATAGGTG TAAATCCTGC GATTGATCT	7980
	GGACCATATT TTTTATTGT ATACAGTAAT TGTGCTGCGA TTATCTCTGT AACGTCTTTC	8040
40	CAATTTGAAC GCACGTGCCC TCCCATACCT CGGGCTTGCT TATATTGTTT GGCTTTGTCT	8100
	TCATTTTCAA CAATAGACGC CCATGCAGCA ACGCGATTAC CATTGTTTTT TTCTAATGCT	8160
	TCAGTCCATA AATCCCAGAG TTTTCCACGA ATATATGGAT ATTTGATTCTG AAGCGGACTG	8220
45	TATTCATACC AAGAGAATGA CGCACCTCGT GGACATCCTC TCGGTTTATA TTCAGGCATA	8280
	TCCGGACCAC AACTTGGATA GTCAGTTTGT TGATTTTCCC AGGTAATCAC ACCATTTTTTC	8340
	ACAAATACTT TCCAAGAACA TGAGCCTGTA CAGTTAACAC CATGTGTTGT TCTTACTTCT	8400
50	TTATCGTGGC TCCAACGTTT TCTGTACATT TTTTCCATT CTCTACTTTT ACTTTCTAGG	8460
	ATCGACCAAT TCCCATTAAT TTTTCTGTT GGCTTAAAGA AATTCAATCC AAATTTTCCC	8520

55

	TAAATGCCC	AAGACTATTG	CTTTAATTAG	ATTGTACATT	TTTTCACAAA	CATAAAATAT	8640
	TAGGGAATCA	CCTAATTACT	TAAGGAATTT	CCCTATCAAT	AACGGGATTT	CATTGAAATA	8700
5	ATACACAATC	ATGTATGGTC	ATGCTTATTG	CCAATCTAAA	TCGTTCAAAT	TTGGCACAAC	8760
	GACAAATAAG	GCTTCAACAC	GAATATATTC	TCTCGGTTGA	AACCTTACTT	ATTCAATTAT	8820
	TTTTTATAAA	TTAGTGACAT	AACACTGTAT	TAGCATCTGC	ACGATCGGTT	GAAATATATG	8880
10	TTACATTTTC	TTGCTGCTTA	ATAAATGCAT	CATAGTAATC	ATATTGCGAC	GAATGATATG	8940
	TGCCATTCTG	TGTATCATTT	GGGTTTAGCA	AACAGCCATA	ACCTTCGTCA	TATAAATGTT	9000
	CACAGAGCAT	AAGGGCGTCA	TGTTTAGAAC	CACTTACTAC	ATAAAATTGC	TTCATAGGAT	9060
15	CATATGATTT	AGGAGTGTTT	TCAGTATAAT	CAACAAC TTC	CCCTATAATA	CATATACCTG	9120
	GTTTCGCCTC	AATTGAATAG	TGTTGCAATT	TTGAAATAAT	ATTACTTAAA	CGCCCCCTAA	9180
20	CAACAACTC	GTTAAAACAC	GATGCTTGAA	AGACAATCGC	TATCGGGTAA	TCAATATCTG	9240
	TGTATTGTTG	TATCTGTGTG	ATAATTTTCC	CTAAACGTTT	TACCCCCATA	TAAATTGCTA	9300
	ACGTGCCACC	ATCTACTAAG	GAATTGACAT	CCACTTCATT	TTCTTCTGAA	TCTTTAAAGT	9360
25	GACCTGTAGA	AAATGTCACA	CTTTTAGCAA	CTGTACGCAT	TGTCAAACCT	GTCTGCATAG	9420
	TAGCAACTGc	tGCGCTCGCT	GATGTCACCC	CTGGTACAAT	TTCAAACGCA	ATATGATGTT	9480
	CATTTAGTAT	GTCGACTTCT	TCTTGCACAC	GACCAAATAT	CGCTGGATCG	CCACCTTTAA	9540
30	GTCTAACAAc	CTTGTTATAT	CGACGCGCTG	CTTCCACGAT	ACAGTCATTT	ATTTTTTCTT	9600
	GCTGAATATG	TTTTGCATAC	GGCTTTTTTAC	CAACATCGAT	AATTTTCAGTA	GTCAAATTTCG	9660
	CATATTGTAA	AATTAACGGA	TTCACTAATC	GATCATATAG	AATGACATCC	gCTTCACGTA	9720
35	TTAAACGCTC	AGCCTTTTTT	GTCAAATAAT	TCGGATTACC	TGGACCCGCA	CCTATCAAGT	9780
	AAACCTTGCC	ATATTCCTCT	ACAGACATAT	ATATACGTTT	CCGTCTGTAA	CTTCTACCTC	9840
40	ATAAACATCT	ACACAACCTT	CATCAGGTTT	TTGAACAATA	CCTGTATTTA	AATCAATTTT	9900
	TTGATCGTGG	AGCGGGCAAA	ATACATATTC	CCCACTCACT	GTCCCTTCAG	ACAATGGTCC	9960
	TTGTTTGTGT	GGACAGATAT	TGTGAATCGC	ATGAATTTTG	CCACTTTCTG	TTAAAAACAA	10020
45	CCCTACCTCT	TTGCCCTTGA	CAATAACCTT	TTTTCCAATT	AGGGGTGTTA	ATTCATCTAT	10080
	AGTTGTCACT	TTAATTTTTT	CTTTTGT TTC	CATGTATTAC	ACCTTCTCCA	CTTCAAAAAT	10140
	TCTACGTGCT	TGAGCATTGC	TAGTTATTGC	TTCCAAGGT	TCAGCTTCGA	CTGCTTTTTT	10200
50	AGCATCCATA	ATGCGTTCAA	ATAGTTCATT	TTGTCTTTCT	GGGTCAAGTA	AGACTTCTTT	10260
	TACATTTTCA	AATCCAAGTC	TTCTTAACCA	TGGCGCTGTT	CTTTCAGCAT	ATATACCTGT	10320
55							

	AGTTGTAAAA AATTCAGCTT TTTCAACTTC TGTACCACCA TTACCACCGA TATAGATTTG	10440
	GAATCCATTT TCAACTGAGA TAATACCAAA ATCTTTAACA CCTGATTCAA CACAACTTCT	10500
5	TGGGCAGCCT GATACACCCA TTTTGAATTT ATGAGGTGTA TCGATGTATT CAAATGTTTT	10560
	TTCTAAACGA ATGCCAAGTC GTGTCGTGTA TTGCGTACCA AATCGACAAA ACTCTTTACC	10620
	AACACAGCTT TTAAGTGAAGC GTGTTTTCTT ACCATAAGCT GATGCTGAAC GCATACCTAG	10680
10	GTCTTCCCAT ATATTTGGTA ATTCTTCTTT TTTAACTCCA TACAAACCAA CACGTTGTGA	10740
	ACCTGTCACT TTAAGTAGTG GCACATGATA TTTCTTAGCC ACTTCTCCTA GACGAATCAG	10800
15	TTGGTCTGCA TCTGTAACAC CCCACGCGAT TTGAGGTATA ACAGAAAATG TACCATCATT	10860
	TTGAATATTC GCATGGTAAC GTTCGTTAGC AAATCTTGAT TCTCTTTCAT CTTTCATGATC	10920
	ATGTGGATAA ACCATGTTTA AATAATAGTT GATTGCTGGT CGACATTTTG GACATCCACC	10980
20	TTTATTTTTA AAGTTTAAAA CATGTCGAAC TTCTTTAGAT GTTTTTTAAAC CTTTCGCTCT	11040
	TATTTGCGTT ACTATTTGAT CGCGTGTCAA ATCAGTACAA CCACATATAC CAGCAGGTTT	11100
	TGCGGCAACA AAGTCATCTC CTAAGGTGTG CTGCAATATT TGAGCAATTT GCGGTTTACA	11160
25	TTTACCACAT GAATTCCTCG CTTTTGTTTT AGCCGTTACT TCTTCAACTG TTGTAAAGCC	11220
	ATTTTCCGTA ATCGCATTTA CTATAGTACC TTTATCAACA CCATTACAAC CACAAATTGT	11280
	TTCATCATCA GCCATATCAG CAATTGATAG CGATGCCTCT TCTCCACCTT TAGTAAGCAA	11340
30	TGATACAAGT GTGTAATCTT CAGTGGATTG ACCTTTTTTC ATCATGTTAT AAAAGCGTGA	11400
	ACCATCATCG ATATCACCAT ATAGTACTGC ACCAACTACA TTACCGTCTT TTAATAAGAT	11460
35	TTTTTTATAG TTATTATCAA CACTATTAAA TATTTCAATA CCTTTAATTT CTGCATTTTC	11520
	TACAATTTGA CCAGCACTAT ACAAGTCACA CCCAGAACT TTTAATGACG TAAATGTTGT	11580
	TGATCCCTTG TATCCGTTTC TTTCTTTATT TGTTAAATGA TCAGCTAATA CTTTACCTTG	11640
40	TTCATATAGT GGTGCAACGA GTCCATAAAC TTTGCCGTTA TGTTCGTCAC ATTCACCAAC	11700
	TGCATATACA TTGCTATCAC TTGTTTGTCAT CACATCATTG ACAACAATAC CACGATTAAC	11760
	ATCTAGACCT GATTCTTTGG CTAATTCTGT GTATGGTCGT ATACCTACTG CCATAACAAC	11820
45	TAAGTCTGCC GGAATCTCGC GTCCATCAGC CAATTTAACA CCCTCAACAT CATCTTCTCC	11880
	TAAGATTTCA GTTGTGTTGG CTTGCATTTT AACTTCATA CCTTGCTTTT CTAGATCTGC	11940
	TTTAAGCATA TTTCCAGCTT TACGGTCTAG TTGCATTTCC ATCAACCATT CAGCTAAATG	12000
50	TAACACCGTT ACTTCATAC CTTGATCTAA TAAACCACGT GCACACTCTA AACCTAGTAA	12060
	TCCTCCACCA ATTACAATTG CTTTCTTTTT AGTCTTAGCA ATGTTTCATCA TTTGTTTCAGT	12120

55

	GAATGCTTTA GAACCTGTG CAAAAATCAA TTTATCGTAT GATACTTCAA TACCATTGTC	12240
	AGTAGTAAGT GATTGATTG CTCTATCTAC TTCAATTACA GGATCATTG TAATTAAGTC	12300
5	GATACCATGT TCCTCATACC ACTCATATGG ATTCATAATT GTTCTTCAA CTGTCATTTT	12360
	ATTTTGTAAG ATATTTGAAA GCATGATGCG GTTATAGTTT GGATAAGGTT CTTTACCTAT	12420
	TACCGTAATA TCATATAAAT CGTTGGCGCG CTCTAATATT TCTTCGATTG TTCGAATGCC	12480
10	CGCCATACCG TTACCAATCA TTACTAGTTT TTGCTTTGCC ATAAAATATG CCCCTTTACT	12540
	CCATAATATT TATTTCAAAA AAAGGTATTA ATTTTTCGTT AGTGCTTTTA TATTTTCATT	12600
15	GGAATCATTG AGCTTTCTAA TCTATCGTTA ATGATTGCTT TTAATAATTG GTCGAAGTTA	12660
	ATTGAAGGTG TGAAGTGTAT ATCTGTATTA ATAACCATGT CATTCAATTG CTGCTTCACT	12720
	TTGTTAACAA GTCTTCCGTC ATATAAAAAT AATGGTACGA CAATCAATTT TTGATACCGT	12780
20	TTGAGATGCT TTTCTAAATC ATGTGTAAAA CTAATCTCTC CATATAGCGT TCTCGCATAT	12840
	GTCGGCTTGC TAATTTGCAA ATTTTGAGCG CATATTGTA ACTCTTCGTG TGCCTTAGTA	12900
	AACTTTCCAT TAATATTGCC GTGTGCAACA ACCATAACTC CAACTTGTG TTCGTCACCT	12960
25	GCTAATGCGT CACAAATACG TTGTTCAATT AATCGTCTCA TTAAAGGATG TGTGCCAAGT	13020
	GGCTCGCTTA CTCTACCTT TATGTCTGGA TACCGTCGTT TCATTTTCATG AACGATATTC	13080
	GGTATATCCT TGAGATAATG CATTGCACTA AAGATTAGCA ATGGTACAAT TTTAAAATGG	13140
30	TCAACCCAC TTTGAATCaA CGTCGTCaTT ACCGTCTCTA AATCCtGATG CTCACTTTCT	13200
	AAAAACGCAA TATCATAGTG ATGTATATCA TCTTTTACTA ATTCAGAAAT AAATGCTTCT	13260
35	AACGCTTGaT TCTGTGCTCC GTGCCTCATG CCAATGTGCA CAATGATATT CCCATTACCA	13320
	TTTACCAACC CTTTCACACG TATTGTATAC CAAATCATTT TGTTTTGTG AAAAGAATCA	13380
	CATTATAATG TAAAATCAGG GAATTCCCTG ATGCCTGTAG TCATGCATAT TCCTTATACA	13440
40	TTTCCCTTT TTGTTAAATC AAAAAAAGCG ACCGATATAT GAATCCCTAC TCAACATTTA	13500
	TTTGAGCAAG CATTAATATA TCGGTCGCTT GTAGTGTATA TTATTATCTT AAAATGGTGG	13560
	TTGGCCTAAT ATTGTTTCGT CAAAGCGCTC GGGTATCAAT ACTTTGCGCA TGATCACACC	13620
45	TAAATCGCCA TCATCATTTT CATGTTGCT GTATATTTCA TAACCTCTT TTTTATAAAT	13680
	TTTAAGTAAC CACGGATGCA ATCTTGCAGA TGTACCTAAA GTAACGCCG CTGACTTTAA	13740
	CGTATCTCGC AAAAATGCTT CTTCAACATA AGTAAGTAAT TGGCTACCAT AGCCTTTCCC	13800
50	TTTACTCTCA GGATTGTGCG CAAACCACCA GACAAAAGGA TAACCCGAAA TACTTTTCAC	13860
	ACTTCCCCAA GGATATCTAA CCGTAATCGT AGATATAATT TCATCATCAA TTGTCATGAC	13920
55		

CCAATCAATA CCTAGTTCTC TTAGAgGCGT AAATGCTTCA TGCATGAGTT CTTGCAATTT 14040  
 TTCTGCATCT T 14051

5 (2) INFORMATION FOR SEQ ID NO: 104:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1885 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

15 TAATCCTCAA CTTnGATTAT ATGGCTTGGG CGCATATGAA CTGCTTAGTT TAGTGTATGA 60  
 CATTCATACA GTTCGCATGA CTATCATACA ACCTCGAATA GATAACTTTT CTAAGAAGA 120  
 20 GTTACCAATC TCAAGATTAC TTCAATGGGG AACCGATTTT GTTAAACCCT TAGCCAGACT 180  
 TGCTTATAAC GGTGAAGGTG AGTTTAAAGC AGGTAGTCAT TGTAGATTCT GTAAGATAAA 240  
 GCATTTCATGT AGAACACGTG CAGAATACAT GCAAAATGTG CCTCAAAAGC CACCACATTT 300  
 25 GTTGAGTGAT GAAGAGATTG CAGAACTTTT ATATAAACTG CCTGATATCA AAAAATGGGC 360  
 TGATGAAGTA GAGAAATATG CGTTAGAACA AGCGAAAGAG AATGATAAAA CGTATCCAGG 420  
 TTGGAAGCTA GTCACGGGAC GTTCAAGGAG AGTGATAACT GATACAAAAG CAGTCCGAGA 480  
 30 CAGGTTAGTT GAAGCGGGTT ATAAACCTGA AGATATTACA GAAACCAAGT TACTTAGCAT 540  
 TACGAATTTA GAAAAATTAA TCGGCAAAAA AGCATTTTCT AAAATTGCAG AAGGCTTTAT 600  
 AGAAAAGCCG CAAGGTAAAT TAACACTTGC TACCGAGTCT GATAAACGAC CAGCTATAAA 660  
 35 GCAATCTGCT GAAGATGATT TTGACAACT ATAAAAATTA AAAAGGACGG TATATAAACA 720  
 TGAAAGCAAA AGTATTAAAT AAACTAAAG TGATTACAGG AAAAGTAAGA GCATCATATG 780  
 40 CACaTATTTT TGaACCTCAC AGTATGCAAG AAGGGCAAGA AGCAAAGTAT TCAATCAGTT 840  
 TAATCATTCC TaAATCAGAT ACAAGTACGA TAAAAGCCAT TGAACAAGCT ATAGAAGCTG 900  
 CTAAAGAAGA AGGAAAAGTT AGTAAGTTTG GAGGCAAAGT TCCTGCAAAT CTGAACTTC 960  
 45 CATTACGTGA TGGAGATACT GAAAGAGAAG ATGATGTGAA TTATCAAGAC GCTTATTTTA 1020  
 TTAACGCATC AAGCAAACAA GCACCTGGTA TTATTGACCA AAACAAAATT AGATTAACGG 1080  
 ATTCTGGAAC TATTGTAAGT GGTGACTATA TTAGAGCTTC AATCAATTTA TTTCCATTCA 1140  
 50 ACACAAATGG TAATAAGGGT ATCGCAGTTG GATTGAACAA CATTCAACTT GTAGAAAAAG 1200  
 GCGAACCTCT TGGCGGTGCA AGTGCAGCAG AAGATGATTT TGATGAATTA GACACTGATG 1260

55



TTGAGGTGTC AAGAATTTGA AATTTATGAA TATAGATATT GAAACATACA GCAGTAACGA 1380  
 TATTTTCGAAA TGTGGTGCCT ATAAATACAC AGAAGCTGAA GATTTTCGAAA TTTTAATTAT 1440  
 5 AGCTTATTCG ATAGATGGTG GAGCGATTAG TGCATTGAC ATGACTAAAG TAGATAATGA 1500  
 GCCTTTCCAC GCTGATTATG AGACGTTTAA AATTGCTCTA TTTGACCCTG CTGTAAAAAA 1560  
 GTATGCATTG AATGCTAATT TCGAAAGAAC TTGTCTTGCT AAACATTTTA ATAAACAGAT 1620  
 10 GCCACCTGAA GAATGGATTT GCACAATGGT TAATTCAATG CGTATTGGCT TACCTGCTTC 1680  
 GCTTGATAAA GTTGAGAAG TTTTAAGACT ACAAAGCCAA AAAGATAAAG CAGGTAAAAA 1740  
 TTTAATTCGT TATTTCTCTA TACCTTGTA ACCAACAAAA GTTAATGGAG GAAGAACrAG 1800  
 15 AAACCTACCT GAACATGATC TTGAAAAATG GCAACAATTT ATAGATTaCT GTATTCGAGA 1860  
 TGTAGAAGTA GAAATGGCGA TTGCT 1885

20 (2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2656 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

30 TAATCCTTAG TTTACTGnCA AATTTCAAAA CACCAGTTCC CTCTATCTGC ATCCATAGAA 60  
 ACTGnATGTT TGTGTCAATA ACCGGATTAT ATTGTGATGn TGTTTGTAAC TCGATTAAAGT 120  
 TATCATCTTT CGAAAAATTA TCTACTACCA TTATTCAACC ACCTTTCCTT CGAATAAACT 180  
 35 CCATTTACCA ACKCCACCAG TACCAAAGTT TCTAACTAAA AATTGATGTG CAGACGGGAA 240  
 GTTATTACGT CTTAATACTT GTGTGTATT ACCTGGTGTA TTCGATTTTA CTTCTAATAT 300  
 CCAACCTGCA ATACCTTTAA AGTCTTTAGG AAAATCAGTA AATCGGTTTG ATTCTTCAGT 360  
 40 AGTGATATAG AAATCTAAAC CAACGATTTT TAAATCTGAT AATTTTGTA TACTCTTAGG 420  
 GATATGTTCC CAATAACCGG CGTTTTCGGG GCAGAAATTC CATGCTCCGT TGTTTTCTT 480  
 45 ATTGAAAATG TCAATGACAC GTTCGAATTT AAGCATATTT CTACCTGTGC TGTTTCTGGT 540  
 AAGTACTTGT CTTAGAGCAC CATTATAGTG TCCAGGCAGT ACATCCAAGA ACCACCCTGC 600  
 ATCTCTAAAC GCTTTCGGTA ACGGGAAATC TAATGCATTT TGTGTGTCTT GaCGTATAGA 660  
 50 TATAGTAATG ACCAACTTCC GTAATATCAC TTAGATATGC TGGGTCTCTG ATTGGTAACG 720  
 GTTTAACACG TCCGCCTGAA TCAGTCATTG ATACTTGAGG TCGATGTTT TTCAAGAATT 780

55

TAGTTACCCC GATTAGAAGT GCTTTACGTC CTGTTTCTAG ATCGTAATAC ATATCTAGAC 900  
 CCTCAGCCTC TTGGAATCT CTTTAAAGT TGTATTTCAC ACCGCCTATA TCGATGCGAC 960  
 5 GTTTAAATAA CAATTCTTTC GTTTTGATAT CGAAGCCTTG TAAGTAGTTA GGGTTGGCTG 1020  
 TATTCGAATC ACCTGTATAC CAATATAAGA TACCTGCATC ATAAGTGATA CCTTGCATAG 1080  
 GTTGTGTATC TGAAGTGTAT TCCATAGGTA TATCCATTTG ATACAATACT TTGTCTATAC 1140  
 10 CTTTATCAAT ATCGTCAGCA CTTCTAACCT CAACAAAGTT CAACGAATTC TTAAGTTGTC 1200  
 TTTCAGTGGG TTTATATTCA CGTCTAAAAA TCATTAAATT TTCTACCGGA TTATAAATCG 1260  
 CTGACGTATA TCTGTCGTTA AATATATTCTG GCATGACATC TTGCATTTCA TTACCATAAG 1320  
 15 TTATTTCTCC AGTTCTATAT TGGAAACGTA CAAACTTGTT GTTTTTGTTA CTGTCCAATA 1380  
 CAGCTGAATA AATCCATAAT TCTCCATCAA TGTATCTATA CGCATTGTGT GTACCGTGAC 1440  
 20 CGCCGTTTTT AACAGCAAT CTATCAATAA ATTGTCCGTT GGGCTTCAAT CTAGATAACA 1500  
 TGTAATGATT ACCTGGACGA GCTTGCCTCA TATAAATAAT TTTCGTTCTA GGGTCTACCC 1560  
 AAAATGATTG CATTACTGCA TTTGTATATG GCGATAAATC AGTGATAAAT TCCGGTTCTT 1620  
 25 GCTCTTTTGG TTCGAATCGG TATTCTGTCTG CTCGATATTC TTTATAGTGT TCATCTACAG 1680  
 CTTTCTCAAC CTTTTTAGTG AAAACATCTA GTGTTGAATA ATCATGATAC AAACGATCTT 1740  
 GCAATGTCTT ATGACCATAA CCTGTATTAT CAACGCGCGC GTCTTTTACT TCGTTGATAC 1800  
 30 CGTCGCCGTT ATGACCTAGT ACCATGTTGC TAAATCGACC GTTTAAATAT GTTAAAAAGT 1860  
 CAGAGACGTT ACTTGTAACA TTTAAATGTT CATACTTTAT TTGTTCTCCA TCATGTGCGA 1920  
 ATACCTCTTT ATTTCTGTGG TATTCAAGAG AGAAATTAAAT ATCCGTCAGC ATGTCTGAAA 1980  
 35 TAAGTTTAAA GTTATACTCA TTTTCATCTA CATATCTGTA GTCAAAGACT CTACTTAAAT 2040  
 CTGTAATTAG TTTATTACTC ATGTTTTCCT CCTTACTAT CCATAAACT GATmATAATT 2100  
 TTTAATAAGC TCATACATAA TAACTTCATG ACCTCTTTCA TTAGGATGTA ATCCATCAGG 2160  
 40 CATGCTAGAT TTTCTAAATG CTGGATTATA TGGTTTGAAA TAATCTGTGT GATAAGCATC 2220  
 ATATACTGGT ACATCCAATT CACTACAAGC CAATATCTGA GCATTGACAT AATCCTCTAA 2280  
 45 AGTTAACCTT AGTTTGTTTT TGTCCGTATC TTTACGGCGT ATCGTTGTAC CACTCATAGG 2340  
 GCATTGCCTA GTAGCTGTCA TTACAAGTAT TTTGAAGCT GGATTATTTT TCCTGATAAC 2400  
 TTCAATTGCA GAACAAAAGG CGCCGTAAAA CGTTTTAGTG TCGGTTTTAT CAGTGCCTAT 2460  
 50 CGGTACGCCT GCCCAATAAC CATGTAACCA GTCATCATCT GTACCTTGTA ATATGATTAG 2520  
 GTCTCCTCTT ATTTGCTCTG CTTGTCTaTA AATGCTGTTT TCTaCCGCTT CTTTACCTAT 2580

55

CTTGCCTAAC ATTTCT

2656

(2) INFORMATION FOR SEQ ID NO: 106:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4854 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

15	AAAATGAGGG TTCTAGCGGA AATTACCAAA AGCGTGGTTC ATACTATGGG CAGCGTAATC	60
	GTATTTCAAA AGAAAAACA CCTAAATGGT TAGaAAATAG AGATAAACCT AGTGAAGAAG	120
	ATTGCGCTAA AGATAATAGC GTAGATGATC AACAAATTAGA GCAAGATCGA CAAGCATTTTC	180
20	TAGATAAATT ATCTAAAAA TGGGAGGAGG ACAGTCAATA ATGAAGCAAT TTAAAAGTAT	240
	AATTAACACG TCGCAGGACT TTGAAAAAG AATAGAAAAG ATAAAnCAGA AGTAATCAAT	300
	GACCCAGATG TTAAGCAATT TTTGGAAGCG CATCGAGCTG AATTmACGAA TGCTATGATT	360
25	GATGAAGACT TAAATGTGTT ACAAGAGTAT AAAGATCAAC AAAAACATTA TGACGGTCAT	420
	AAATTTGCTG ATTGTCCAAA TTTCGTAAAG GGGCATGTGC CTGAGTTATA TGTTGATAAT	480
	AACCGAATTA AAATACGCTA TTTACAATGC CCATGTAAAA TCAAGTACGA CGAAGAACGC	540
30	TTTGAAGCTG AGCTAATTAC ATCTCATCAT ATGCAACGAG ATACTTTAAA TGCCAAATTG	600
	AAAGATATTT ATATGAATCA TCGAGACCGT CTTGATGTAG CTATGGCAGC AGATGATATT	660
	TGTACAGCAA TAACTAATGG GGAACAAGTG AAAGGCCTTT ACCTTTATGG TCCATTTGGG	720
35	ACAGGTAAAT CTTTTATTCT AGGTGCAATT GCGAATCAGC TCAAATCTAA GAAGGTACGT	780
	TCGACaATTA TTTATTTACC GGAATTTATT AGAACATTAA AAGGTGGCTT TAAAGATGGT	840
40	TCTTTTGAAG AGAAATTACA TCGCGTAAGA GAAGCAAACA TTTTAATGCT TGATGATATT	900
	GGGGCTGAAG AAGTGA CTCC ATGGGTGAGA GATGAGGTAA TTGGACCTTT GCTACATTAT	960
	CGAATGGTTC ATGAATTACC AACATTCTTT AGTTCTAATT TTGACTATAG TGAATTGGAA	1020
45	CATCATTTAG CGATGACTCG TGATGGTGAA GAGAAGACTA AAGCAGCACG TATTATTGAA	1080
	CGTGTCAAAT CTTTGTCAAC ACCATACTTT TTATCAGGAG AAAATTTTCAG AAACAATTGA	1140
	ATTTTAAAAT GATTGGTGTA TAATGAATAC AAATCTAAAT CGTTTAAATG ATTGAAGACA	1200
50	AGATGATCTA ATCAATATTA CACAGAAAGC CATTGTTTGA TGAGAATATG GTTAATAAAT	1260
	TAGATGATTA CTA CTTCATT TATGGTATTT GTAATGAATA CCCGGATCAA GACCGTTATC	1320

55

	CTCGTCCCTT GTATAGGGGC GGGATTTTTT GTTTTTTCA GACATAAATG TTTGTTGGTG	1440
	TCATAAATTC CCTGTTTATT GTTAATAGGT TTAATGTTAA AACGATGATT GTTGTTC AAT	1500
5	TTTTTAACGA GGT CAGATAA AAGTATTTAT AAAGCAAATA GGAGGGTTTA ACATGGAACA	1560
	AATTAATATT CAATTTCCAG ATGGTAATAA AAAGGCGTTT GATAAAGGTA CTA CTA CTGA	1620
	AGATATAGCA CAATCAATTA GTCCTGGATT ACGTAAAAAA GCTGTTGCCG GCAAATTTAA	1680
10	CGGGCAACTT GTAGATT TAA CTAAACCGCT TGAAACTGAT GGATCAATTG AAATTGTGAC	1740
	ACCAGGTAGT GAAGAagcGT TAGAGGTATT ACGTCATTCT ACTGCACATT TAATGGCACA	1800
	CGCGATTAAA AGGTTATATG GTAATGTTAA ATTTGGTGTA GGTCTGTAA TAGAAGGTGG	1860
15	ATTCTACTAT GACTTCGACA TTGACCAAAA CATCTCATCT GATGACTTTG AACAAATTGA	1920
	AAAAACAATG AAACAAATCG TTAACGAAAA TATGAAAATC GAACGAAAAG TGGTTTCACG	1980
	AGATGAAGTG AAAGAGTTAT TCAGCAATGA TGAATACAAA TTAGAATTAA TCGACGCGAT	2040
	TCCTGAAGAT GAAAATGTAA CATTATATAG TCAAGGTGAT TTTACTGATT TATGTCGTGG	2100
	AGTTCACGTT CCATCAACAG CTAAAATTAA AGAGTTTAA CTATTATCTA CAGCAGGTGC	2160
25	ATACTGGCGT GGAGATAGTA ACAACAAAAT GTTACAACGT ATATACGGTA CTGCTTTCTT	2220
	TGATAAAAAA GAATTGAAAG CACATTTACA AATGTTAGAA GAGCGTAAAG AACGTGATCA	2280
	TCGTAAAATT GGTAAAGAGT TAGAACTATT CACAAATAGC CAATTAGTTG GTGCTGGTTT	2340
30	GCCATTATGG TTACCTAACG GTGCAACAAT TAGACGTGAA ATTGAACGTT ACATTGTTGA	2400
	TAAAGAAGTT AGCATGGGAT ATGACCACGT TTATACACCA GTACTTGCTA ATGTTGATTT	2460
	ATACAAAACA TCTGGTCACT GGGATCACTA TCAAGAAGAT ATGTTCCAC CAATGCAGTT	2520
35	AGATGAAACT GAATCTATGG TATTACGTCC AATGAACTGT CCACATCATA TGATGATTTA	2580
	TGCGAATAAA CCACATTCAT ATCGTGAATT ACCTATCCGT ATCGCTGAGC TAGGAACGAT	2640
	GCATAGATAT GAAGCAAGTG GTGCTGTATC AGGATTACAA CGTGTTCTGTG GTATGACTTT	2700
40	AAATGATTCA CATATCTTTG TTCGACCTGA TCAAATTAAA GAAGAATTCA AACGCGTTGT	2760
	AAACATGATT ATTGATGTGT ATAAAGACTT TGGTTTCGAG GATTATAGCT TTAGATTAAG	2820
45	TTATAGAGAC CCTGAAGATA AAGAAAAGTA CTTTGATGAT GATGATATGT GGAATAAAGC	2880
	TGAAAATATG CTAAAGAGG CAGCGGATGA GCTTGGCTTA TCGTACGAnG AAgCGATTGG	2940
	TGAAGCGGCA TTCTATGGTC CGAAACTAGA TGTTCAAGTT AAAACAGCGA TGGGTAAAGA	3000
50	AGAGACATTA TCAACAGCAC AACTTGATTT CTTATTACCA GAACGTTTTG ATT TAACTTA	3060
	TATTGGTCAA GATGGTGAAC ATCATCGTCC AGTTGTTATT CATCGTGGTG TTGTATCAAC	3120
55		

EP 0 786 519 A2

	AGCGCCAAAA CAAGTTCAAA TCATTCCAGT TAACGTTGAT TTACATTATG ATTATGCGCG	3240
	CCAATTACAA GATGAATTGA AATCTCAAGG CGTTCGTGTA AGTATTGATG ACCGTAATGA	3300
5	AAAAATGGGT TATAAAATCA GAGAAGCTCA AATGCAAAAA ATACCTTATC AAATCGTAGT	3360
	TGGGGATAAG GAAGTTGAAA ATAATCAAGT GAATGTGCGT CAATATGGAT CGCAAGACCA	3420
	AGAAACAGTT GAAAAAGATG AATTTATCTG GAATCTAGTT GATGAAATTC GTTTGAAAAA	3480
10	ACATAGATAG ACAGTTGTCG CAATAAAATG CTTTAAACT TTTATTGCGT ATCAAGTTT	3540
	ACAGGGTTGA TTATGCGTGA TGAATCCTGT ATATTACAAG TTAGTTAAAA TATTAAATTG	3600
	AGTTAGAGGT TGCATGTTTA ATTAGTAACT TGTCAGAAGT ATTTATGGTA CATAAGTTGA	3660
15	ACAAGTGAAG GGTAAAGATG CCGAAATAGA TATAAACCAT AAATTATATC TATTGGGACA	3720
	GTTTTCGAAT AGGAACTGTA CTGTCACAGA ATGTGATGTG CTACCTTATA TAGATAATTG	3780
20	CCAAAGTGGT TGCATATCTT AAAGGTATGT AGCCACTTTT TTACTTTTAA TATCACTATG	3840
	TTCTGTAAAA AAGGGTATGA AAGTGAATAA AGGTTATTTA TTTCTTGGCC TCTAAAACAT	3900
	GGAAAGGGAG CTTATATGTC AAAAGTTCAA AATGAAAGTA ACAATGTTGT CAAAAGGGGA	3960
25	CTTAAAGATC GTCATATTTT TATGATTGCG ATTGGGGGTT GTATTGGTAC AGGTTTATTT	4020
	GTAACCTCTG GTGGAGCAAT TCATGATGCA GGTGCTTTGG GTGCATTAAT AGGATACGCA	4080
	ATTATCGGAA TAATGGTATT TTTCTTAATG ACGTCACTTG GCGAAATGGC TACGTATTTG	4140
30	CCAGTATCAG GTTCATTTAG TACATATGCT ACAAGATTG TTGATCCATC TTTAGGGTTT	4200
	GCGCTTGGTT GGAACCTATT GTTTAACTGG GTAGTGACTG TAGCAGCAGA TATTACGATT	4260
	GCAGCACAAG TCATTCAATA TTGGACACCA TTGCAAGGCA TACCCGCTTG GGCATGGAGT	4320
35	GCGTTGTTCT TAGTTATAAT TTTTAGTCTG AATTCGTTAT CAGTTCGCGT CTATGGTGAA	4380
	AGTGAACTACT GGTGGCATT GATAAAAGTG GTTACAGTTA TTGTTTTCAT TGCAATTGGT	4440
40	TTATTAACGA TTGTCGGAAT CATGGGTGGT CATGTTGTAG GATTCGAAAT ATTTAATAAA	4500
	GGTGAAGGTC CAATTCTTGG TGGCAACTTA GGAGGAAGTT TGTATCAAT TCTAGGTGTA	4560
	TTCTTAATCG CTGGTTTCTC ATTCCAAGGT ACTGAGTTAA TTGGTATTAC GGCTGGTGAA	4620
45	TCAGAAAATC CTGAACGTGC TGTGCCGAAA GCAATTAAAC AAGTATTCTG GAGAATTTTA	4680
	TTATTTTACA TTTTAGCCAT TTTTGTATC GGTATGTTAA TTCCTTATGA TAGTAGTGCA	4740
	TTAATGGGGG GTAGTGATAA TGTAGCAACG TCTCCATTCA CATTAGTGTT TAAAAATGCT	4800
50	GGATTTGCGT TTGAGCATC ATTTATGAAT GCAGTCATTT TAACGTCTGT GTTA	4854

(2) INFORMATION FOR SEQ ID NO: 107:

55

(A) LENGTH: 2488 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

10	ATCAAAAATT GATTGTTTTC nATTTTTTGT TTCAGCGCGG GATCTTTTAC GTCTTTTG TG	60
	AAAACGaTTT TATTATTAAC TACTTTTACT GGATAACTTT TGTATGTCGA GTCAGTAGCA	120
	TTTTTCTAT CGTTTGTAGT TGTGTCATAT TCACCAgTTA TTTTATGTGT GTTCTTATCT	180
15	ACCTTTAACA ACATACGGTC TTCTTTTAAA AGCTCATCTG ATCCAACAAC TGAATAAGAG	240
	GATTCTATAT ACCATGTGTC TTGATCATTa TTTTCATAAT GGGGATTATC GTGACCATCA	300
	ATTCATAAA GCGTTTCTAA GTTTTTAATA GGATACGTAC TTAGTACTTT TTTAAGACCA	360
20	TCTTCAAAT GAATTTGTTC CCACTTCATT GCCAAAAACA TATCGCCACT GACTACAATT	420
	GAAATAATAA TAATTGCTGC TAAGTTTAAC CAGAAAATTT TATGTGCTTT CATAcATTCC	480
	CACCGTTTCT CAAAATACTT CATTAACTACT ATAATAATAT ATTTTGAAAA ATATTTACAT	540
25	CAGTATTAAA GTGAATATCA AATTTTAAAT TTATGAAAAT AATAGATATT TATAAAAAGC	600
	GGAAAAGAGA TACAATAAAA AACTGCATGA CGTTTGAGAC GTCACACAGT GTAACTAAAA	660
30	ATTTAAAAAG TTGTTGCTAA TTTTTCAGCA TTATTAATAC TAGTTGCTTT AATTTCTTCA	720
	GTCTTATGAG GTTCAGCATT GTGTCCTTCA ATAATGATTG TTTcATATGA TGGCACACCT	780
	AAGAATGTCA TAATTGTTCT TAAATAACGG TCACCCATTT CAAAATCAGC AGCAGGTCCT	840
35	TCAGTATAAT ATCCACCACG TGATTGAATG TGTAATACTT TTTTGTCAGT TAGTAAACCT	900
	TGTGGTCCTT CAGCAGAATA TTTAAAAGTT TTACCTGCAA TTGAAATAGC ATCAATATAT	960
	GCTTTAACTA CAGGTGGGAA AGAAAGGTTT CACATAGGCG TTACAAATAC ATATTTATCT	1020
40	GCACTTAAAA ATTCTTCTAA AATGTCACCT AATCTTGAAA CTTTCATTG TTCATCATCA	1080
	GTTAACGTTT CGCCATTACT CATTTTTCCC CAACCAGTTA ATACATCTTT GTCAATAACT	1140
	GGAATATAAG TTTCArATAA ATCAATATGT TTCACTTCAT CATCAGGATG TTGTTGTTGA	1200
45	TATGTTTCGA TAAATGCTTT ACCAGCCGCC ATAGAATTTG ATACCAGTTC ATTAAAAGGG	1260
	TGTGCTGTAA TATATAATAC TTTTGCCATT TGAAAATTCT CCTCTGkTTC TGTTATTTTC	1320
	TTAAGTATAA TTATTATACT CGATATAAAA TTTAATATCA ATCAAAATAT TCAAATTACC	1380
50	ATCATTTTCT TCATCTATAT nTGGCAGTAC TACTAAAGTA TGAGTGcATT TAATTATGaa	1440
	ATAGTTGATT TaGAATaTAT ACTTAATACC CAAAATATAT GAAGGATGGA TGCCACTATG	1500

55

ATTATTTATA TAGATGACAT TCAAAAATGG TTTAACCAAT ATACCGATAA ATTGACACAA 1620  
 AATCATAAAG GACAAGGACA CTCAAAATGG GAAGACTTTT TTAGAGGGAG TCGGATTACT 1680  
 5 GAGACTTTTG GTAAATATCA ACATTCACCA TTTGATGGTA AGCATTATGG CATTGATTTT 1740  
 GCATTGCCAA AAGGTACACC AATTAAAGCG CCGACGAATG GTAAAGTAAC ACGTATCTTT 1800  
 AATAATGAAT TGGGCGGCAA GGTATTACAG ATTGCCGAAG ACAATGGAGA ATATCACCAG 1860  
 10 TGGTATCTAC ACTTAGACAA ATATAATGTC AAAGTAGGTG ATCGAGTCAA AGCAGGTGAT 1920  
 ATTATTGCAT ATTCAGGCAA TACAGGTATA CAAACGACAG GCGCACATTT ACATTTTCAA 1980  
 15 AGAATGAAGG GTGGCGTAGG TAATGCATAT GCAGAAGATC CAAAACCGTT TATCGATCAG 2040  
 TTACCTGATG GGGAACGTAG CCTATATGAT TTGTAGTTAT AGAAGGGTGC CCGCAGTCTA 2100  
 AAAAATTAAG CAATCATTGT GTGAGTATGA TACTTACATA ATGGTTGCTT TTTTCAATGA 2160  
 20 AAATCGTAAT GCTAAGTCAT ACTTGTGTTGA TTTAGATATT ACTTAAAATG TAAGACAAGG 2220  
 TTGTTAGCAT TGGCAGTGAA ATATCGCACA TAAAAACAT TATTGTCACA CTAGAAAATA 2280  
 GTTGTGCACT ATATCAATTT TCTGTATAAA AGTTTAATTC TGACAGTAAT GTAAACGTTT 2340  
 25 ACAATTTATG ATTGACATTA ATAATGACTG AATATATGAT TTATGTAAGT ATTTGTGCAA 2400  
 CGTTTTTACA AAGTGTATTG CACaAyCAAA CTGTAAACaA aGTATGGGGg GCCATAACAT 2460  
 30 GGCAGAACTA AGTTAGAGCh TATTAAAA 2488

## (2) INFORMATION FOR SEQ ID NO: 108:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4093 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

TTTTCTTTAT TTCAAmCTGT ATATTaATGA TGTCACCTTCA TTTGATACGA TTCTTGATAA 60  
 CCTATTCAAA ATTCCGCCAA ATAACATAAA TATTATATAA ATGCCGATAC TTTTAATCAT 120  
 45 TTTCTACTTT TTCTTCGATA CGGAAACTTG TTTTCGAATT GAACACTTCA CCAGCTTTTA 180  
 AAATTGACGG TGCTTTTTTCA CCATATAAAT TAATATCATT TGGTAAAAAT TGTGTTTCTA 240  
 50 ATGTAAAGCC AGAATGTGGT TTATAAATAT TAAATGGACT ATCCCACTCA TCAGGCTGGT 300  
 TAAAAGTAAA GAACACAACA TGAGGCATAT CTGTATCGAC CTCTAACATA AATTCATGAT 360  
 TTTCAACATA CATTTTATGT TCACCAACTG TAAATGGGTG ATCGAGACCA CCAAAACGTG 420  
 55

	TATCTTCAAA CACTTCATGT AAATCTAGAA TATCACCTGT AACAATATTT CGCTCATCTA	540
	ATACATACAT ATCTAATTGA TTACTTGAAA TGCGATGATT ATCAACGACA TTATTATCTC	600
5	GATTCAAATT GAAGTACACA TGATTGCTAG GACTAAACAA TGTGTCTTCT GATGCAACTG	660
	CTTCGTATTC AATCGACCAT TGGTGATCCG CATCATAAAT ATGTGTAATC GTCACATCGA	720
10	TATCACCCGG GAAATGATCA TCAGCTGATT TCAACACCGT CTAAATATA ACTTTAATTT	780
	GAGCAATTTT ATTTCTAATT TCATAATCAA ATAACCTTATT GTCCAAACCA TGACATCCAC	840
	CATGTAAATG ATGTTACCCG TTGTTTTTTT CTAAGTATA TTCTTTACCT TTCAACTTAA	900
15	ATTTAGCATT ATCAATTCTA CCGCTATATC TTCCTATAGA AGCACCAAAT TTAAGAGGAT	960
	TACTATGATa AAATTCATCC GCTTCAACAA CATTTCCAAG AACAATATTA TTATCATGAT	1020
	ATTTCCAAGA CACTACTCTT GCTCCATAAT TCGTAAAAAT AATTTTAGTT TCATCATTAT	1080
20	CAATTTTGAT TAAATCTACA CCTTGTCTTT GGTGCTCAAC TTCAACTATC ATTTTACTT	1140
	CTCCCTTCTA ACCACAAGTG TTCAAGCTCT GCTGGGTAGC AACATTACTA AAACACCTAC	1200
25	AATACAAATG ATTGCACCGA TAACATCATA TTTATCTGGC ATTTGTTTAT CTACGACCAT	1260
	CGCAAAAATC AAATCATGA TGATAAATAC GCCACCATAT GCTGCATATA CTCTCCGAA	1320
	TGATGGAAAT GATTGAAATG TCGCAATGAC ACCATATAAC ATGAGTATCG CACCGCCTAT	1380
30	TAGCCCAACA AGTGAAGACT GTCCTTCCCT AAGCCACAGC CAAATCAGGT ATCCCCCACC	1440
	TATTTACAT AAGCCAGCTA ATATAAATAT AAAAATCGGA TATAACATGA AATCACTCCA	1500
	TCACACATTT GCTATCAATA ATCTATCGGC TACATATCAT TTGTTTACAT TTCTTCTTAC	1560
35	TTACATTCC CATTTTAAAA AGTTCGTTTT CACATTCATA TTGTACACTT TTTTAGACAT	1620
	TATTCTATAG CTAAATATAA AAAAATAAGA GTAACACGCT TTCATCATCA TTTTATATGA	1680
40	TAAATGTGTG TCACTCTCAT CAATTTTATT TTTTAAATAC ACGTTTCATT GAATTAAATA	1740
	AGCCACGTTT AAATGTAAGT ACTGAATCTT TATATGTTTT AATTGCAATC CATATCAAGA	1800
	CAGTACCAT TACAATTGAG ATTAAAGAAC TTAAGATGAC CTCATATATT TGAAGCCCTG	1860
45	AAGTTTGAGC GCGTACAAC TATTGAAATG GCGCTAAAAA CGGAATATAA CTTGTGATTA	1920
	AAGCAAGTTG TCCATCAGGA TTATTTATCG TGAATATCGC GATATAAAAT GCAATCATAC	1980
	CAAGTAATGT CAGTGGCATC AAAGATTGAT TTAAATCTTC TATTCTAGAT GTTAATGATC	2040
50	CGAGGATGGC TGCAAGTAAT ACATACGCCG TAATTCCAAC AATACTACTT ATAATTCCGA	2100
	CAATAATAAT TTGCCAAGAC AATTGATTCA TTTCCACGTT AAAACCTTGT AGCAAGTCTT	2160
55	TTAAGTCAAA GGCAAAAATG CATATAACTG CCATCAATAC AATTAAAATA ATCTGAGTCA	2220



	TAATAATCAT	TTCAATGACA	CGCGATGTTT	TCTCACTAGC	AATTTCCATA	GCTATTTGAG	2340
	ATGCATAATT	TAAAACAATG	AAGAACATTA	GAAAGATAAT	GCCATmaGcT	AAAGCATAGT	2400
5	TGAAAATCTT	TTGTCCTTCT	GATACTTTAT	CGACTTCATC	ATTAGAAATC	ACCTTATTAT	2460
	CAACTTTACT	TTGTGCTTGT	AATTTTTGTG	AGTCTTCTTT	GTTGATATTT	AATTCCCCGG	2520
	CTACCATATT	TGTTTGAATA	GCTGTAAGCA	GTGCTTGTA	TTTTTGTGAA	TCTTCATGAC	2580
10	TTACTCGCTT	CTCACTAATG	ATTGTCCCTT	GTAACGTGCG	ATTTTGATT	ACCTTGATAA	2640
	TATAAGCTTT	ATCAAGTTTA	TGTTTTTTTA	CTTCTTTTTC	AGCATCTTCT	ATAGAACTT	2700
15	TAGTAACTT	AGCATCACTA	TGAAATGTAT	TCGCCTGTTG	CTTGAAAACC	TTATAGATTT	2760
	GTTCAATCGG	TGCTGCTACA	CCAATTTTAT	CTGGACCATC	ATCAAACATG	TTAATAATCT	2820
	TATCAATGTT	AGATAGGCCA	ATCATTAAGG	CAGCAATAAT	AATCATAAAA	ATTACAAATG	2880
20	ATTTAGCTTT	AATTTTTTTG	ATATATGTCA	AAGTAAATGT	CGCCCCAAAC	TTATGCATCC	2940
	TTGCCACCAA	CCTTCTCAAT	GAATATATCT	TGTAATGATG	GTTCTACAAC	TTGGAATCGT	3000
	TTAACATAAC	CTTGATGTGC	CACAACTTGA	TAAATATCTT	TGGCTACGTC	TTCATTCTCA	3060
25	ATCGTCAACT	GAAGACCTTG	CTTCATGTTT	TCATATGAA	TGATGCCTCT	AATGTTTGTT	3120
	AAATCTGGTA	GTGTTGTTTC	TGATTCAATG	ACAACTTTCT	TGTTACCATT	AGATGCACGT	3180
	ACATGATTGA	TATCACCAGA	AACAACAAGT	TGACCTTTAT	CTAAAATACA	AACATCATCA	3240
30	CATAATCTT	CAACATGCTC	CATACGGTGA	GAACATATAA	CGATTGTACT	GCCCCAATCA	3300
	TTTAAGTCTT	TAACTGCTTC	TTTTAATAAC	TCAACATTAA	CTGGGTCTAG	ACCACTGAAA	3360
35	GGCTCATCTA	ATATTAGTAA	TTCTGGTTTA	TGTAACATAC	TTGCTAACAG	CTGAATTTTT	3420
	TGTTGATTCC	CTTTTGATAG	ACTATCAATT	CGTTTTTTGC	GGTTTTTCAGT	AATATCAAAA	3480
	CGCTCAAGCC	AATACGATAT	TTGCTGTTGT	ATTTCTGTTT	TTGACATTCC	CTTTAAAGTT	3540
40	GCCAAATATT	TCAATTCTTC	TTCAACTGTC	AATTTCCCAT	GTAAACCGCG	TTCTTCCGGT	3600
	AAATAACCAA	TACGATTGTA	CATTGTTTTA	TCTAGTTTTT	TACCGTTATA	CGTrTGTGT	3660
	CCTTCAGTTG	GTTCACTTAA	GCCTAAAATC	ATACGAAATG	TCGTTGTTTT	ACmTGCACCA	3720
45	TTTCTTCCTA	GAAAACCTAA	CATTTTACCT	GATTCTAACT	TTAATGAAAT	ATCATTTACT	3780
	GCCGTCATCT	TGCCAAAACG	TTTCGTAACA	TGTTCAATTA	CAAGTCCCAT	ACTTTGCCTC	3840
50	CTAAAAAAT	ATGTATTTAT	CTTAATATAA	CATTTCCATT	CTCTATAAAT	GCAATATTTT	3900
	TAAATGAAT	TTATTTTTTAA	AATTTCTGAA	ATTGAAAAAT	TTAAATAGTG	CCATTTTTGC	3960
	ATGTTAAGTA	TCATTAGCAC	TAGATATGTT	TTTCCATGC	CTTTATTGCC	TTATTTGTAA	4020
55							

CTTnCCGGTG TTT

4093

## (2) INFORMATION FOR SEQ ID NO: 109:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17846 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

15	TGCCAACTA CCTTTTGACA GTCGTTGCTG TACTTCAGGA TGATCAATCA CATATnTTAC	60
	TTTATCAAAT AGGGCATCTT CATCATTTTT AGTAATTAAA TAACCATTGA AATCTGAAGT	120
	AATCAGTTCG TTAGGTCCAT ATTTAATATC ATAACTAATA ACTGGAACAC CATGTGCTAA	180
20	AGATTCAAGT AGCGCTAAAG AGAAACCTTC CATGTTACTT GTTATTAAAC TCAAATAGGC	240
	ATCGCTATAT TCTTGGTCTA GATTGCTTAA AAAGCCGCGT AAGTAAACAT GATTTTCCAA	300
	TCCATATTTT TGTATCAATT CATTTAATTT TTTACTTTCA GaaCCAAAAC CATACATATG	360
25	AaGCTCTATT TTTGGGACAT ACGATACTAA GCGTTTAATT AATTCAATTT GTTGATGTAA	420
	TTGTTTTTCA GGTGAATAAC GAGCAACGGA AATTAATTTA AACTGCGCT GATCTAATGT	480
	TTGGACTGGT GTATCAATTG TTTCACTATA GCCGACAGGA ATATTAACAA CTGGAATAGT	540
30	ATGGTTAATA CGTTTTTCAA CATCTAATTT TTGCTGCTCA GTAGAAACGA TAATTGCACG	600
	ATATCGAGAT AAATTTTCAA ACATCGCTTT ATATACATTT TTAAATGGCG ATGAATCTAA	660
35	TGCATCAATA TTTTAAATGT GTGTACTGTG AAGCACAGCT ACTACTGGA TTGACTCAGG	720
	CGTTAAGTTG AAAATAGGTG CTGTGTACAC ATTACGATCA CTGAAAATA AATCCCCATG	780
	TTGATATAGT TGTTTAATGA AAAATGCGCC TAATTCCGTT TCATTATTAA AGAAATATTG	840
40	TTTGTTAGCA TAGTAAACAA TAATTTTTTG TACTTCTGGT TTGCCATCCT TGTAAGAAAA	900
	ATACTTTTCT AATTTTGTGT CACCTTCTGG ATTATAGAAA AATTCACATA ATGTTTGTG	960
	TTTATCAACA AGAATCCTAC TACAACCTAA AAAGCCACGC ACATCATAAA AATCACGTTT	1020
45	TACTTtTCGT CTTTGACTAT CAAAATGATT TACATAATCT AATATACGAT ATTTAGGATC	1080
	TTGAAAATGG GCATACATTA AGAAACGCTC TTGATCATAT ATTCTAAAGT CATGACTATT	1140
50	TTCAACATGT TTTAAAGTAT AATGACATTC ATCAGTCCAA TACGACAACC AGTCAAATGG	1200
	TTCAATTGCGT TCTAAATATG TTGCTTCTTG GAAGAAATCA TACATATTAA TATAGTCAGA	1260
	ACTAGTAATA TAATTTTGGG CATTTCTATA TAAATATCTA TTCCATGACA GAAATACACA	1320

	CCCAGTTAAA TTAACACCTA AACTATTACC TACAAAATAA TTCATTTACA ACACCACTTA	1440
	TATCTATTTT TTATAATTAT ATCACATAAT ATTTAATTAC TTCTTTTAAC TGGAAGATGT	1500
5	GTTTATTTAT AAAACAACAA ATTTTGATAT TTATAATGAT AGTAGTTATT CAATCACTAC	1560
	GACCCaATAT ATCATkGTAG AGCTTAGGAT ATTGATTTAT GACTCAGGCA CATCAAATGa	1620
	GAgGATTTAT AAArGAGATA TACAACCTA GAAGGTATAA TAAAAACGCG CAACTAATGT	1680
10	TACGCGTTTG AATTAATCAT ATGATATTAT TTGCGATACT TTAATTTAGC GAAAgcATCA	1740
	TGTTGATGGA TAGACTCTTC ATTACGACAT TCGATATCGA AACCGTCTAA CCAATCAAAT	1800
15	TCAACTAAGT CCGCGGCAAT TAAACGAATT AAGTCTTCGA CAAAACGTGG ATTTTCATAT	1860
	GCACGCTCTG TCACACGTTT TTCATCAGGA CGTTTTAAAA TAGGGTATAG AATTGAACTT	1920
	GCAITAGCTT CCATTGCATC TAAAATTTTA TTTTATAGT CATCAACTAT GTCTTGATCT	1980
20	TTATTAATAT ATGTTTAAAC AGTGACAACA CCACGTGGT TGTGCGCTGA ATACTCACTT	2040
	ATTTCTTTTG AACAAAGGGCA TAGCGTTGTG ACAGTTGCTT CAATAGTAAG TTCTTTACGT	2100
	GTAnCTTTAT CACCGTCAAT TGCTAATCCA TAAGTGACAT CGGCATTACC AACTGCTTTA	2160
25	ATATTTGTGG TTGGACTATA GCGATCAAAG AACCATTTC CAGAAACATC AACGCCTGCC	2220
	GCAITTTGTT TCATATTCGT TTGTAAAGTG CGTAACACCT GATAAAGTGT ATTAAATTCA	2280
30	AGTTCAATAC CATTATCATA GTGCTTTTCA ACACTTTCGA TTATACGGCT CATATTAATA	2340
	CCTTTTTCGT CTTTGTAA ACTTGTGAA AAATAAATG TGCCAGCTGT TTGATACTGG	2400
	TCAACAAGTA CAGGGTACAC TAAGTTTTTA ATACCAACTT CTTCTATTTC AAATAAAAAA	2460
35	TCTTTATGTG TACTTTGTAA ATCTGTCATT TCGTTCCTAG TAGTAGGTTT CGTGCCTTCA	2520
	ATAGGATCTA CGGAACCAAA GTGTTTCCAA CGACCTTCTC GTGTCGATAA ATCAAATTCA	2580
	GTCA <sup>7</sup> TTTTT TCCTCCGTTA AGATTTAAAG TGATATGTCC AATATGGTTC GACTGTAA	2640
40	AAGCTGTGTT GTTTACCATC GATTTCAAGG CTTGCTAATT GTTTTAAAAA TGGACCTGTT	2700
	TGAGAAGCAT GTGCTTCAAA TGCCTTAATT TTAAGTTCTT TAAAATCTGT AATATCATTT	2760
	TGAATATCAG GTTCTCCAAG AGCTTCGGTT GCATCATTAC TGAACGCAAC TAAAGTTAAA	2820
45	CGAGGGCGTT CTTCTTTAGG CATGCGTTCA ACCGTTGAA TTACAGCGTC TGCTGTTGCT	2880
	TCGTGATCAG GATGTACTGC ATATCCAGGA TAAAATGAAA TAATCAATGA TGGATTGTA	2940
50	TCATCGATTA AAGATTTAAT CATACCATCT ATATGTTTAT AGGGTTCAAA TTCGACAGTT	3000
	TTGTACGTA AACCCATTTT TCTTAAATCA GTAATACCGA TAACTTTACA AGCTTCTTCT	3060
	AGTTCACGCT CACGAATACT TGGAATGAT TCGCGTGTG CAAATGGGGG ATTACCTAAA	3120
55		

	TAATTTGCTA ATGTGCCTGC AGATGAGAAG GTTTCATCAT CAGGATGTGG AAATATTACT	3240
	AATACATGTC TTTCGTCAGT CATGTTGATG CCTCCTCTAT AAATTAAATG GTCGCTCACT	3300
5	AATTTGAAGT GCTGCAGCGA GTTGACCTTC GTAATTAAAA CCTGCAATTA AAAATTCATC	3360
	ATGCTCATTG ACCTCAAAAT GCGTTAGACC TTGTACATAA ACCCAACCAC CATTTGATAG	3420
10	TTTAAGACCA ATGCGATAAG GTTCTTTATT ACCACCTTTT AGTTGTGCAT GCGTATATGT	3480
	TATTTGTATG TTTCTTAAAA AAGTACCAGC ATTAAAAACA CGTTGATCGA AATGGTTCGC	3540
	ATAGGCCCCA TTTGTCGTTT CAACATGCAG ATACACAGGT TTATGTTCAA AAGAAGCAAG	3600
15	TAAATCTATA ACTTCTTGTT CTTTAATTGG TTCCAACACG TTCACTCCTT ACACTATCAA	3660
	TGTGTTTATC TTTCTATTTT ACTAAAACT ATTCGATAAT TGTATACGAT TGCTCAATTA	3720
	TTTATAAATT AATTTTCATG AAGGGTAATT ACTCAGGATT ACGTAATCAT ACAGCATTAG	3780
20	TTTTTTACTT TTA AAAATCA AAAATTTGTT GGAATTTGAA AAGTGTTAAA CATTAAAAAT	3840
	GATGCTATAT TAATGGTGTA TGAATGAATT CATAAGTTTT TAAATGTAT TAAATTTGTG	3900
	GAGGCATGTA AACAAAGAAA GTATTAACT TAGGATCGAA AAAACAAGCA TCATTCTATG	3960
25	TTGCATGTGA GTTATATAAA GAGATGGCAT TTAATCAGCA CTGTAACTA GGTTTAGCAA	4020
	CTGGTGGTAC AATGACAGAT TTGTATGAAC AACTTGTTAA GTTGTTAAAT AAAAATCAGT	4080
30	TAAACGTAGA CAATGTATCC ACGTTTAATT TAGACGAATA TGTAGGTTTA ACCGCATCAC	4140
	ATCCGCAAAG TTATCACTAT TATATGGATG ACATGCTTTT CAAACAATAT CCTTATTTTA	4200
	ATAGAAAGAA CATTCATATT CCAAATGGAG ATGCCGATGA TATGAATGCG GAAGCGTgCA	4260
35	AAATATAATG ACGTTTTAGA ACAACAAGGT CAACGTGATA TTCAAATTTT AGGTATTGGT	4320
	GAAAATGGTC ATATTGGATT TAATGAACCT GGTACGCCGT TTGATAGCGT TACTCATATC	4380
	GTTGATTTGA CTGAAaGTAC TATTAAGGCT AATAGTCGAT ATTTTAAAAA CGAaGATGAT	4440
40	GTTCCAAAGC AAGCCATTTT GATGGGACTT GCTAATATTC TTCAAGCCAA ACGTATCATT	4500
	TTACTCGCAT TTGGTGAAAA GAAACGTGCT GCTATTACAC ATTTATTAAA TCAGGAAATT	4560
	TCTGTTGATG TTCCAGCCAC ATTACTTCAC AAACACCCGA ATGTTGAGAT ATATTAGAC	4620
45	GACGAAGCTT GCCCGAAAAA TGTTCGAAA ATTCAATGTCG ATGAAATGGA TTGATTGCAA	4680
	TGTTTAATTA AGAAATGCCT CGGGAAAGGT TCCAATAGAA AGATAAAAAAG CATTGGAAGG	4740
50	ATGATTTTTA GTGGAATTAC AATTAGCAAT TGATTTATTA AACAAAGAAG ACGCGGCTGA	4800
	GTTAGCAAAT AAAGTAAAAG ATTATGTAGA TATCGTAGAA ATCGGTACGC CAATCATTTA	4860
	CAACGaAGGT TTACCAGCAG TTAAACATAT GGCAGACAAC ATTAGTAATG TAAAAGTATT	4920
55		

	CGCGGATGTA ATTACAATAC TAGGTGTTGC AGAAGATGCA TCAATTAAAG CAGCTATTGA	5040
	AGAAGCTCAT AAAAATAATA AACCAATTACT AGTTGATATG ATTGCTGTTC AAGATTTAGA	5100
5	AAAACGTGCA AAAGAACTAG ATGAAATGGG TGCTGATTAT ATTGCAGTAC ACACTGGTTA	5160
	TGATTTACAA GCAGAAGGGC AATCACCATT AGAAAGTTTA AGAACCGTTA AATCTGTTAT	5220
	TAAAAATTCT AAAGTTGCAG TAGCAGGTGG AATTAAACCA GATACAATTA AAGATATTGT	5280
10	CGCTGAAAGT CCTGATCTTG TTATTGTTGG TGGCGGAATC GCAAATGCAG ATGATCCAGT	5340
	AGAAGCTGCG AAACAATGTC GCGCTGCAAT CGAAGGTAAG TAATATGGCT AAATTTAGTG	5400
15	ACTATCAATT AATTCTAGAT GAATTAAAGA TGACTTTGTC ACATGTTGAA GCGGATGAGT	5460
	TTTCAACTTT TGCATCCAAA ATACTACATG CTGAACATAT ATTTGTAGCT GGCAAAGGAC	5520
	GTTCAGGATT CGTGGCGAAT AGTTTTGCAA TGCGCTTAAA TCAGCTCGGC AAACAGGCAC	5580
20	ATGTTGTTGG AGAATCAACG ACACCTGCGA TTAAGTCGAA TGATGTATTT GTAATTATCT	5640
	CTGGTTCAGG TTCCACGGAA CATTTAAGAT TATTAGCAGA CAAAGCAAAA TCAGTAGGTG	5700
	CTGACATCGT ATTAATTACT ACAAATAAAG ATTCTGCAAT AGGCAATCTA GCTGGGACGA	5760
25	ACATCGTTTT GCCTGCAGGT ACAAATATG ATGAACAAGG CTCGGCACAA CCATTAGGAA	5820
	GTTTGTTTGA ACAAGCATCT CAATTATTTT TAGATAGTGT TGTAATGGGA TTGATGACTG	5880
	AAATGAATGT TACGGAACAA ACGATGCAAC AAAATCATGC TAATTTAGAA TAAAATAAAG	5940
30	ATAGTCGATA ATATGATGCC TAGGCAGAAA TATTATCGAT TATTTTTTTA TTTAAATAAT	6000
	AAATTATAGT ATAATATCAA TAATAACGA ATAGGGGTGT TAATATTGAA GTTTGACAAT	6060
35	TATATTTTTG ATTTTGATGG TACGTTGGCA GACACGAAAA AATGTGGTGA AGTAGCAACA	6120
	CAAAGTGCAT TTAAAGCATG TGGCTTAACG GAACCATCAT CTAAAGAAAT AACGCATTAT	6180
	ATGGSAAATAC CTATTGAAGA ATCATTTTTA AAATTAGCAG ACCGACCATT AGATGAAGCA	6240
40	GCATTAGCAA AGTTAATCGA TACATTTAGA CATACATATC AATCTATTGA AAAGGACTAT	6300
	ATTTATGAAT TTGCGGGTAT AACTGAAGCC ATTACAAGTT TGTATAACCA AGGGAAAAAA	6360
	CTTTTCGTGG TGTCTAGTAA GAAGAGTGAT GTATTAGAAA GAAATTTATC GGCTATTGGA	6420
45	TTAAATCACT TGATTACCGA AGCTGTTGGA TCCGATCAAG TAAGTGCATA TAAACCAAAT	6480
	CCTGAAGGCA TACACACAAT TGTGCAACGC TACAATTTAA ATAGCCAACA AACGGTGTAT	6540
50	ATTGGTGATT CAACGTTTGA TGTTGAGATG GCACAACGTG CTGGTATGCA ATCTGCAGCT	6600
	GTCACCTGGG GTGCACATGA TGCAAGGTCA TTACTTCATT CAAATCCGGA TTTTATTATT	6660
	AATGATCCAT CAGAAATTAA TACCGTATTA TAAACTTGT TAAACAGAG AATACCATGG	6720
55		

	ATTTAAAATA AATATTTATT AAACATTATG AATTTTAAAA GAGTAATGTC TGA	6840
	CTCTCGTTG ATAATTTATT TTTGTAAAAA TAAATTAAAG TAATGACAAA GTTATTGAAG TAAATTGAGT	6900
5	ATAAACATTT AAATACGATG TCGAAAATGG CGATAGCATA TCACTTACAT GAAGTTGTGT	6960
	GctATCGCTA TTTTtagTTA TAATTCCAAA AAGTTAATCG TTCGATGATT TAAGAATTAT	7020
	TATTGTTTAA TTCAAATGTA TGAGGGTATA AAATCATTGA ATTTAATTCG ATAAAGCGAA	7080
10	ATTTTTGAAC AAACATACTT TTGTATTTAT ATAAAAGTTT AAATCTTAT AAATTTGACA	7140
	AAACTAATTA ACTCCGTATA ATTATGAAAC ATACAAGAGG GAGTGTATGA ATTCATGGAT	7200
15	TTTAATAAAG AGAATATTAA CATGGTGGAT GCAAAGAAAG CTAAAAAAC CGTTGTTGCA	7260
	ACCGGTATCG GTAATGCAAT GGAATGGTTC GATTTTGGTG TCTATGCATA TAcAACTGCG	7320
	TACATTGGAG CGAACTTCTT CTCTCCAGTA GAGAATGCAG ACATTCGACA AATGTTGACT	7380
20	TTCGCAGCAT TAGCCATTGC GTTTTTATTA AGACCAATTG GTGGTGTCGT ATTTGGTATT	7440
	ATTGGTGACA AATATGGACG TAAAGTTGTA TTAACATCTA CAATTATTTT AATGGCATT	7500
	TCAACATTAA CCATTGGATT ATTGCCAAGC TATGATCAAA TTGGACTTTG GGCACCAATA	7560
25	CTATTATTGC TTGCAAGAGT ACTACAAGGG TTTTCAACAG GTGGAGAGTA TCGGGGGGCA	7620
	ATGACATATG TTGCCGAATC ATCTCCAGAT AAGCGTCGTA ACTCATTAGG TAGTGACTA	7680
	GAAATTGGGA CATTATCAGG TTACATAGCT GCTTCAATTA TGATTGCTGT ATTAACATTC	7740
30	TTTTTAACAG ATGAACAAAT GGCATCATTT GGTGGAGAA TCCCATTCTT ACTCGGTTTA	7800
	TTCTAGGAT TATTCGGCTT ATATTACGT CGTAAGCTGG AAGAATCACC AGTTTTCGAA	7860
35	AATGATGTTG CAACACAACC AGAAAGAGAT AACATTAACT TTTTACAAAT CATCAGATTT	7920
	TATTACAAAG ATATATTTGT ATGTTTGTGTA GCTGTTGTAT TCTTCaATGT TACAACTAT	7980
	ATGGTAACTG CATATTTACC AACCTATTTA GAACAAGTTA TTAAATTAGA TGCAACGACA	8040
40	ACAAGTGTAT TAATTACTTG TGTATGGCA ATAATGATTC CATTAGCATT AATGTTTGGT	8100
	AAGTTAGCGG ATAAAATAGG TGAAAAGAAA GTATTTCTAA TTGGTACTGG TGGGCTAACA	8160
	TTATTCAGTA TCATCGCATT TATGTTATTA CATTACAAT CATTTGTTGT AATAGTAATC	8220
45	GGTATATTTA TATTAGGATT TTTCTTATCA ACTTACGAAG CGACAATGCC AGGGTCGTTA	8280
	CCAACGATGT TTTACAGTCA TATAAGATAT CGAACTTTAT CAGTAACATT TAATATCTCT	8340
50	GTTTCGATAT TTGGTGGTAc GAcGCCATTA GTkGCAMCaT GGTTaGTTAC GAAA	8400
	ACTGGA GATCCATTAG CmCCTGCGTA TTATTTAACA GCAATCAGTG TTATTGGCTT TTTAGTTATT	8460
	ACATCTTAC ATTTAAGTAC AGCAGGAAAA TCTCTAAAAG GTTCGTATCC AAATGTAGAT	8520
55		

	GAACGTAAGA	ATTAGAGATT	TTAATAAAAA	GTATAAATCA	ATCGTATATA	AGCACTTTAA	8640
	AGCTAGTAGG	TTCTGCTAAC	TTTAAAGTGC	TTTTTAAATT	GAGAACTGTA	ATTAGCCGTA	8700
5	ATAAAGTTTT	TGTATATACA	TAAACCCCCA	CTGCAATGAT	TATCGCAATG	GGGGAAAGAG	8760
	GGGACTTAAA	GCATATGTTT	AGCTTTGAAT	ACTTAAATTT	CTCTTGCTAT	TGAAATGTTA	8820
	GGATGTAAAT	ATGTCTTAGA	GTATTTTGTC	CAACGCAATT	AATATTGAGA	CTCTAACCTT	8880
10	CAATATTATT	ATAGAGAACA	CAAACCTAAA	TAGATTGGGT	GACTTATTTG	TGTCAGTTAT	8940
	TGCGATTGCG	ATAACTTCTT	TTCTCTATAT	ACATATAGTA	ACGTCTTATC	TAATAAAAAA	9000
	CATGGTACTA	CAGTATCAAA	TTTATCTAGG	GCTTAAAGTTT	GATTTTTATA	ATAGGCAGGT	9060
15	TTACCTGATA	AAAATACTTA	TTCATTATAT	AATGTTAACA	ATATGTATTT	TAAAGTTTAC	9120
	ATTGAGTGAG	GGATATTGAT	GAACGTAATT	TTAGAACAGT	TGAAAACACA	TACTCAAAAT	9180
20	AAACCTAATG	ACATAGCATT	ACATATCGAT	GATGAAACAA	TTACATATAG	TCAACTAAAT	9240
	GCCCGCATCA	CTAGCGCAGT	TGAATCTTTG	CAGAAATATT	CACTTAACCC	TGTCGTTGCT	9300
	ATTAATATGA	AATCACCGGT	GCAAAGTATT	ATTTGTTATT	TAGCTTTGCA	TCGTTTACAT	9360
25	AAAGTGCCTA	TGATGATGGA	AGGTAAATGG	CAAAGTACTA	TACATCGTCA	ATTGATTGAA	9420
	AAATATGGTA	TTAAAGATGT	AATTGGAGAT	ACAGGTCTCA	TGCAGAATAT	AGACTCACCG	9480
	ATGTTTATTG	ATTCAACGCA	ATTACAGCAC	TACCCCAATT	TATTACATAT	TGGTTTTACT	9540
30	TCAGGGACAA	CTGGACTGCC	AAAAGCATAT	TATCGTGATG	AAGATTCATG	GTTGGCTTCT	9600
	TTTGAAGTTA	ATGAAATGTT	GATGTTAAAA	AATGAAATG	CAATAGCAGC	CCCTGGACCA	9660
	CTATCGCACT	CGTTAACATT	ATATGCGTTA	TGTTTTGCTT	TAAGTTCCGG	TCGTACTTTT	9720
35	ATAGGACAGA	CCACTTTTCA	TCCTGAAAAG	TTACTTAATC	AATGTCATAA	AATATCATCA	9780
	TACAAGTTG	CTATGTTTCT	TGTTCCAACG	ATGATTAAAT	CATTATTGTT	AGTTTACAAC	9840
	AATGAACATA	CAATCCAATC	ATTTTTTAGC	AGTGGAGATA	AGCTGCATTC	TTCTATTTTT	9900
40	AAAAAGATAA	AAAATCAAGC	AAATGACATA	AATTTGATTG	AATTTTTTGG	TACATCGGAA	9960
	ACCAGTTTTA	TCAGCTATAA	CTTGAATCAG	CAAGCACCAG	TTGAATCAGT	AGGTGTGCTA	10020
45	TTTCCAAATG	TGGAATTGAA	AACAACGAAT	CACGATCACA	ATGGTATAGG	AACTATTTGT	10080
	ATAAAAAGTA	ATATGATGTT	TAGTGGCTAT	GTAAGTGAAC	AATGTATAAA	TAATGATGAA	10140
	TGGTTTGTTA	CTAATGATAA	TGGCTATGTA	AAAGAGCAGT	ATTTATATTT	AACGGGACGT	10200
50	CAACAGGATA	TGTTAATTAT	TGGTGGTCAA	AATATATATC	CAGCACATGT	TGAACGCCTT	10260
	TTAACGCAAT	CTTCGAGCAT	TGATGAAGCA	ATTATCATCG	GTATTCCAAA	TGAGCGTTTT	10320
55							

	CAATTTTAA AAAAGAAAGT GAAaCgnTaT GAAATTCCAT CGATGATTCA TCATGTAGAA	10440
	AAGATGTATT ACACTGCAAG tGGTaAAATT GCTAGAGAAA AAATGATGTC GATGTATTTG	10500
5	AGAGGTGAAT TATAATATGA ATCAAGCAGT CATAGTTGCA GCTAAACGAA CTGCATTTGG	10560
	GAAATATGGT GGCACTTTAA AACATTTAGA GCCaGAACAA TTGCTTAAAC CTTTATTCCA	10620
	ACATTTTAAA GAGAAGTATC CAGAGGTAAT ATCTAAAATA GATGATGTAG TTTTAGGTAA	10680
10	TGTTGTTGGG AATGGTGGCA ATATTGCAAG AAAAGCATTG CTTGAAGCGG GGCTTAAAGA	10740
	TTCAATACCT GCGTCACAA TCGATCGGCA ATGTGGGTCT GGACTTGAAA GTGTTCAATA	10800
	TGCATGTCGC ATGATCCAAG CCGGAGCTGG CAAGGTATAT ATTGCAGGTG GTGTTGAAAG	10860
15	TACAAGTCGA GCACCTTGA AAATCAAACG ACCGCATTCT GTGTACGAAA CAGCATTACC	10920
	TGAGTTTTAT GAGCGTGCAT CATTTGCACC TGAAATGAGC GACCCATCAA TGATTCAAGG	10980
20	TGCTGAAAAT GTGGCCAAGA TGTATGATGT TTCAAGAGAA TTACAAGATG AATTTGCTTA	11040
	TCAAGTCAT CAATTGACAG CGGAAAATGT AAAGAATGGA AATATTTCTC AGGAAATATT	11100
	ACCTATAACC GTTAAAGGAG AAATATTCAA CACTGATGAA AGTCTAAAAT CACATATTCC	11160
25	GAAAGATAAC TTTGGCCGAT TTAAGCCCGT GATCAAAGGT GGGACCGTTA CCGCTGCGAA	11220
	TAGTTGTATG AAAAATGATG GTGCAGTTTT ATTGCTTATT ATGGAAAAAG ATATGGCATA	11280
	CGAATTAGGT TTCGAGCATG GTTTATTATT TAAAGATGGT GTTACGGTAG GTGTTGATTC	11340
30	TAATTTTCCT GGCATTGGTC CAGTACCAGC CATTTCCAAC TTAATAAAAA GAAATCAATT	11400
	AACGATAGAA AATATTGAAG TCATTGAAAT TAACGAAGCG TTCAGTGCAC AGGTAGTTGC	11460
	CTGCCAACAA GCTTTAAATA TTTCAAATAC GCAATTAAAT ATATGGGGTG GTGCATTAGC	11520
35	ATCAGGTCAT CCATACGGTG CAAGCGGTGC CCAATTAGTG ACTCGATTAT TTTATATGTT	11580
	TGACAAAGAG ACTATGATTG CATCTATGGG GATAGGGGGA GGTCTAGGAA ATGCAGCATT	11640
40	ATTTACTCGA TTCTAACCAG CGATTAAATG TGTCATTTTT TAAGGATAGT GTGGCTGCAT	11700
	ATTATCAGTG TTTTAACCAA CCTTATAGAA AAGAAGTACC ACCATTAAATG TGTGCGTCAT	11760
	TATGGCCAAA ATTTGATTTA TTTAAAAAAT ATGCAAATAG CGAACTGATT TTAACAAAAT	11820
45	CAGCAATTAA TCAAACCTCAA AAGATAGAAG TAGACACAAT ATATGTAGGG CATTTAGAAG	11880
	ATATTGAATG CCGACAGACT CGCAATATCA CACGTTATAC AATGGCTTTA ACATTAECTA	11940
	AAAATGATCA ACATGTCATA ACGGTtACAC AAACTTTAT TAAGGCGATG AAGTAGAGAT	12000
50	GGAGTTTAAT GAGATATGGA TAAATGAATA TTTGGCGCTC GTAAATGATG ATAATCCAAT	12060
	ACATAATGAG ATTGTGCCAG GACAATTAGT GAGTCAAATG ATGCTGATGG CTATGTCATT	12120

55



	ATTCATTGAA CAACACGAAC ACGAAATTAT AGCAATTAAT GACGATGGAG AGATTAAAAT	12240
	AAAAATTTCT TTGAGCACAA AAAAATAACC GATATTAGCT GCATGAACGC ATATTAATTA	12300
5	GGAGATGAAA GGACAGCTAA TATCAGTTAT GTATTGTTAT TATTATTGGG AACAGAGATG	12360
	AATATAGGTT ACGTTTCTTT CTTTGACCGG GGATGCATTA ATCTAAAATA ATAATAACAA	12420
10	CTATATCAAT GTTTAATAAA TTCTGGATTA TTGGAACGAT TAGTCAATTT AACTAACTTT	12480
	CATATGATCT ATATCGTCTT GTAATAAAGA GAGCAATTTG AATATTTTCTAG TATCACTAAA	12540
	TGAATCGTCA CATTTAATTG AAACATGCTG AAACGTTTGT GTTATAATTT CATAAACTGG	12600
15	TGCGCCTTCA TGGTGATACT GTCGATAAAT AATCATAACC TATATTACCT CCTTTGCTAC	12660
	TCTATGGTTA TATTATAAAT AACATTTTGA TGTGTGACAT CAACCTTAAG TATCAACTTT	12720
	TTATCAGACA TAGAACGTAT GATTTACTAA GACTATTTAT GTATAAAAGT TCTAAATAAA	12780
20	TATATATTTA TAGAGTCGCC TGGCAGTCAT TTGGGAATA TAACATATAT GATTAGAGAG	12840
	GCATCTATCG CAAAAGAATG ATAATGATAG AGGTATTGAG CATATAGATG AGTTTAAGTT	12900
	CATCTTGAAA ATAAAGGGTT ATTTAGTCAT AGATGTAGAT GTATAGGAAA TATTTGTATG	12960
25	TATTGTTTGA TATGTATGAA ATTTTCAATA AAAGCTAATA ACGCTTATAT GTAACTTTCA	13020
	AATTTAAATT ATATACAGAG CATGATGATT ATAAAAAAT AACCACATCA CATAAATTGA	13080
	GTTCATACCC AATTTAAGTG GTGTGGCTAA TAATGTTGAT TTATAGATGA ACCGCCTAAT	13140
30	CGTTAAACCT CTGTTACTTC AACATCGATA TGTTCAATAC GGTGTGATGC ACCGTGATCC	13200
	ACAGGACCAA CAAATCATT CATTTTCCAA CCGTTTTTAA TAGCAGAAGC GACGAAAGCT	13260
35	TTGCGCTAA TCACAGCTTC TTTCGGTGAC TTACCGTTAG CTAAATATGC AGTTGTTGCC	13320
	GCAGCAAATG TACAACCAGC ACCATGGTTA TAACTTTGTT GGAACATGTC TGTTGTTAGT	13380
	TGATAAAATG TTTGACCATC ATAGTATAAG TCATACGATT TATCTTGATC TAAAGCTTTG	13440
40	CCACCTTTAA TGATGACATG CTGTGCGCCT TTATCAAAGA TAATTGTTGC AGCCTTTTTC	13500
	ATATCTTCAA TTGAATTTAA TTTACCTAAT CCTGATAATT GACCCGCTTC AAATAAGTTT	13560
	GGTGTCACTA CCGTTGCTTT AGGTAGTAAA TATTTAATCA TCGCCTCAGT ATTTCCAGGA	13620
45	TTAAGCACTT CATCTTCGCC TTTACAAACC ATGACAGGAT CTACTACAAA ATATTGTGCA	13680
	TTAGATGCCT CATATACTTC TCCAGCACGT TTGATTATCT CCTCAGTACC TAACATACCT	13740
	GTTTTAATAG CATCAGGTCC GATTGATAAA GCCGTTTCAA GTTGTTTTTC AAATACATCC	13800
50	ATTGGTAATG GTGTAACATC GTGTGACCAT GTATCTTTAT CCATAGTAAC GATGGCAGTT	13860
	AAAGCGACCA TGCCATACGT ATCTAATTCT TGGAACGTTT TCAAATCTGC TTGCATACCT	13920
55		

	CACTCCTACA TAATAATATT GTATTCATCA TATCATTTTT AACCTAATTG AAAAATATTA	14040
	AGCATTCAAT ATTTGATGAT TGTGAAATG AATCATTCAT ACTATTGTAA CTTTGTAAAA	14100
5	TGTCATTAC TTTAGATAAG TGTGATATGT TAAAATATGT CCTGAGGTGA GATTGAATGG	14160
	AATGGTCGCA AATTTTTTCAT GACATAACAA CGAAACATGA CTTTAAAGCT ATGCATGATT	14220
	TTTTAGAAAA AGAATATTCG ACTGCAATCG TATACCCTGA TAGGGAAAAT ATATATCAAG	14280
10	CGTTTGATTT AACACCGTTT GAAAATATCA AAGTTGTTAT ATTAGGACAA GACCCGTATC	14340
	ATGGTCCAAA CCAAGCACAT GGATTAGCAT TTTCAAGTGA ACCTAACGCA AAATTCCCTC	14400
15	CATCTTTACG TAATATGTAT AAAGAATTAG CAGATGATAT TGGATGCGTT AGACAAACAC	14460
	CGCATTTACA AGATTGGGCA AGAGAAGGCG TCTTGTTATT GAATACAGTT TTAACCGTAA	14520
	GACAGGGTGA AGCAAATTCT CATCGTGATA TTGGTTGGGA AACATTTACT GATGAAATTA	14580
20	TTAAAGCAGT GTCTGATTAT AAAGAACATG TTGTCTTTAT TTTGTGGGGG AAACCTGCAC	14640
	AGCAAAAAAT AAAGCTTATC GATACATCTA AACATTGTAT TATAAAATCA GTGCATCCTA	14700
	GTCCACTGTC TGCATATAGA GGATTCTTTG GATCAAAACC GTATTCCAAA GCGAATGCCT	14760
25	ATTTAGAGTC AGTAGGAAAA TCACCAATTA ATTGGTGTGA AAGTGAGGCG TAGATGTTGA	14820
	ATAGAGAAAC TTTAATAGCA CGAATTGAGC AAGAATTAGT ACAAGCAGAG CAGGCACAGC	14880
	ATGACCATGA CTTTGAAAAA CATATGTATG CCATACATAT ATTAACATCT TTATATGCTT	14940
30	CAACATCAAA TACACCACAT ATTGGTGAAC AACAAATGAA TCGTCGTATT GCTAACCATA	15000
	ATCAAATGCC ACAATCACAA ATAACGCAGC CAACTCATCA AGTGACAGTT GCTGAAATTG	15060
35	AAGCGATGGG TGGTAAAGTA AATACGCATT CAGCACATCA TCATAATAAG TCATATTCAC	15120
	AACCTTCAAA CCAACAACAA AGATTAGCGA CAGATGATGA CATTGGCAAT GGTGAATCCA	15180
	TATTGATTT TTAAAAAGCA ACAATGAAAC ATAATTACTT AATAGCTTGT TAAGTATGTA	15240
40	GGTTAATAAT CAAGACGCAT ATACTTTTAT TCGAGTGTTC GGATTTAAAC ATTTATTAAT	15300
	ACTGAATTAT ATAAGGAGAG GTAGCAATGA AATTATTTAT TATTTTAGGT GCATTAAACG	15360
	CGATGATGGC TGTCGGTACA GGTGCATTTG GTGCGCATGG TTTACAAGGA AAAATAAGTG	15420
45	ATCACTATTT ATCAGTATGG GAAAAAGCAA CGACGTATCA AATGTACCAT GGCTTAGCAT	15480
	TATTAATTAT AGGTGTAATT AGTGGTACAA CTTCAATCAA TGTTAACTGG GCTGGCTGGT	15540
	TAATATTTGC TGGTATTATT TTCTTTAGTG GATCATTATA TATTTTAGTA TTAAC TCAA	15600
50	TTAAAGTTTT AGGTGCGATT ACGCCAATTG GTGGCGTATT GTTCATCATT GGATGGATAA	15660
	TGTTAATCAT TGCGACATTC AAATTTGCTG GTTAAATTTT AAAACTTTAG ATTACCTATG	15720
55		

TGGGTATAGA ATACCTTCGA GGTGAGTTTT TATTTATGGA AAAAAAGAAT AAGCAAATAG 15840  
 ATAGAGGCCGA TTTAAACAA AACCTATCTG AAAAGTTTGT ATGGGCGATT GCATATGGTT 15900  
 5 CATGTATCGG ATGGGGCGCA TTCATCTTAC CAGGAGACTG GATTAAGCAG TCAGGTCCGA 15960  
 TTGCAGCATC AATTGGTATA GTTATTGGTG CATTATTAAT GATATTAATT GCGGTTAGTT 16020  
 ATGGGCGATT AGTAGAGAGA TTTCCAGTAT CAGGGGGCGC GTTTGCCTTT AGTTTCTTAA 16080  
 10 GTTTCGGCAG ATATGTGAGT TTCTTCTCAT CATGGTTTTT AACTTTTGGT TATGTCTGTG 16140  
 TCGTTGCTTT AAAGCGACC GCATTCAGTT TACTAGTTAA ATTCTTATTG CCAGATGTCT 16200  
 TAAATAATGG GAAACTATAC ACCATTGCGG GCTGGGACGT TTATATTACG GAAATCATTAA 16260  
 15 TTGCGACCGT ATTACTACTT GTATTCATGC TAGTAACGAT TCGTGGCGCA AGTGTATCTG 16320  
 GATCATTACA ATATTATTTT TGTGTGGCGA TGGTAATCGT CGTATTATTG ATGTTCTTTG 16380  
 20 GTTCATTCTT TGGTAATAAT TTTGCACTTG AAAATTTACA ACCGTTAGCT GAACCTAGCA 16440  
 AAGGATGGTT AGTGTCTATT GTGGTTATTG TATCCGTGGC ACCATGGGCA TATGTTGGAT 16500  
 TTGATAATAT TCCACAAACA GCAGAAGAGT TTAACTTTGC ACCAAACAAG ACATTTAAGC 16560  
 25 TTATCGTGTA CAGTTTATTA GCAGCATCAT TAACTTATGT TGTCATGATT TTATACACTG 16620  
 GTTGGTTATC AACAAATCAT CAAAGTTTAA ATGGGCAGTT GTGGTTAACA GGTGCTGTTA 16680  
 CACAAACAGC ATTTGGTTAT ATTGGATTAG GTGTATTAGC AATTGCAATT ATGATGGGTA 16740  
 30 TATTTACTGG TTTAAATGGA TTCTTGATGA GTTCAAGTCG CTTGTTATTT TCTATGGGAC 16800  
 GTTCAGGTAT TATGCCAACA ATGTTTAGTA AATTACATAG TAAATACAAA ACACCATATG 16860  
 TCGCAATCAT ATTCCTAGTA GGAGTGTCGT TAATTGCACC TTGGCTAGGA AGAACTGCAT 16920  
 35 TGACTTGGAT TGTAGATATG TCATCTACTG GTGTATCCAT TGCCTACTTT ATTACATGTT 16980  
 TGTGTCAGC GAAATTATTC AGTTATAACA AACAAAGTAA TACGTATGCA CCGGTTTACA 17040  
 40 AAACGTTTGC TATTATCGGC TCATTTGTAT CATTCAATTT CTTAGCGTTG TTATTAGTGC 17100  
 CAGGTTCTCC TGCAGCACTG ACTGCACCGT CTTATATTGC ATTACTTGA TGGTTAATCA 17160  
 TCGGTTTAAT ATTCTTTGTG ATTCGATATC CTAAATTGAA AAATATGGAT AATGATGAAT 17220  
 45 TAAGTCGCTT GATTTTAAAT AGAAGTGAAA ATGAAGTTGA TGATATGATT GAAGAACCTG 17280  
 AAAAAAGAAA AACTAAATAA TAAAAGAATC GCACAATAAA CCTTCTTCAT TCGGAGGCGT 17340  
 ATCGTGCAT TTTTGTATT ATAAATTGAC ATTTAAGACG AGGCAGCTGA ACCTTATATA 17400  
 50 TAATTGCTAA GAGTTAGGGC TGAGCCATTT CTAACAAATA TTTATAATCG TTTAAAAGAT 17460  
 TTCACGAACC CAGAAACAAT TAATTGGAA ATTTGGTCGG CGAATAATAA ACCTAATGCG 17520

55

AAGACTAAAT TTTTGTAGC ATCGTATGCT AAGCCACCAG GTACTAATGG AATGATACCC 17640  
 GTTACCATAA AAATGATGGC AGGTTCTTTT TGTTTACGAG CCATATAATG ACTTAACAAG 17700  
 5 CCTAATGCTA AACTACCAA GAAACTAGAG TATATAGTGT GCACATTAAA GCCGTTGAAG 17760  
 AATAAGGTGT AAACCATCCA TCCACACGTA CCAACGAAAC CACATGATAG ATATAATTTT 17820  
 CTAGGTGCAT CAAAAATGAC GCAGAA 17846

10

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5544 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

ATTGACACTT GGTGAAAGTA ATATCGCCGC GCTATTTTGG CAAAATGGAC ACTTAGAACC 60  
 TGAGTTACAA GATGAACAGC CAATTAATAT ATTAGGATCT GkTCAAATCA ACGAATGGAA 120  
 25 TGGTAAATCAA TCACCGCAAA TAATTATTCA AGATATTGCG ATGAATGAAC AGCAAATATT 180  
 AGATTATAGA AGTAAGCGAA AAAGTTTACC TTTTACAGAA AATGATGAAA ATATTGTCGT 240  
 GCTTATTCAT CCTAAAAGTG ATAAAGTAAA TGCGAATGAA TATTATTATG GTGAAGAAAT 300  
 30 TAAACAACAA ACTGATAAAG TAGTATTAAG AGATTTACCA ACGTCAATGG AAGACTTGTC 360  
 TAATTCCTTG CAACAACCTGC AATTTTCTCA ACTTTATATA GTTTTGCAAC ATAATCATTC 420  
 GATTTACTTC GATGGTATAC CTAATATGGA TATTTTAAA AAGTGTTATA AAGCATTAAT 480  
 35 AACTAAACAA GAAACAAATA TCCAGAAAGA GGGTATGTTA TTGTGTCAAC ATTTAAGTGT 540  
 GAAACAGAT ACACTTAAAT TCATGTTGAA AGTTTCTTA GACTTAAAT TTGTAACACA 600  
 40 AGAAGATGGT TTAATTCGAA TCAATCAACA ACCTGATAAA AGATCGATTG ATTCCAGCAA 660  
 AGTATATCAA TTAAGACAAC AACGTATGGA TGTTGAAAAG CAATTATTAT ATCAAGATTT 720  
 TTCAGAAATA AAAAATTGGA TAAAGTCACA ATTGTCGTGA GCAATTTAGG AGGAAATATT 780  
 45 AATGGATTTA AAGCAATACG TATCAGAAGT TCAAGATTGG CCGAAACCAG GTGTTAGTTT 840  
 CAAGGATATT ACTACAATTA TGGATAATGG TGAAGCATAT GGCTATGCAA CAGATAAAAT 900  
 TGTAAGATAC GCAAAAGACA GAGATGTTGA TATCGTTGTA GGACCTGAAG CGCGTGGCTT 960  
 50 TATCATGGC TGTCCTGTAG CTTATTCAAT GGGGATTGGC TTTGCACCTG TTAGAAAAGA 1020  
 AGGGAAATTA CCTCGTGmAG TCATTCGTTA TGAGTATGAC CTAGAATATG GTACAAATGT 1080

55

	ATTAGCTACT GGTGGTACGA TTGAAGCAGC AATAAAATTA GTTGAAAAAT TAGGCGGTAT	1200
	CGTAGTAGGT ATTGCATTTA TAATTGAATT GAAATATTTA AATGGTATTG AAAAAATTAA	1260
5	AGATTACGAT GTTATGAGTT TAATCTCATA CGACGAATAA TAAATAATAT AATTTTATCA	1320
	AATGAAATCC TTCATCAAT GTATAAGAAC CAATGACTTA ATTAAAAAAG TTGTTTAAGT	1380
	TTTCTTAACA TGAGATGTTA GGATTTTTTA TTTACTGAAA ATGTTAGATG ATTGAGCATT	1440
10	ATACCTTAAT AACATCGTTT ATTTATTTC A TAAATTGTAG TATCATAGAA CTAATATTTA	1500
	AAAAATGAAA CAGTAGATTT AGGTCGAATT TTTGTAAAAG TTTTAAAAGT AGGAATAGTA	1560
15	TACAAATTAA ACTCGCTCAA GTAAAATTAA TATTACGATT AATGACGACA GGATAAATAT	1620
	TTATCGTCGA CGGACGTATG ATTGGTGTGG GACAAATACT ATTCAACAAG AGTACCTAAA	1680
	TCATTGTTTA AGGCGAAGTA ATAAATATGA ATGGGGTGTA TCATATAATG AACAAACGAAT	1740
20	ATCCATATAG TGCAGACGAA TCTTCACAA AGCAAAATCA TATTTGTCAG CAGATGAATA	1800
	TGAGTATGTT TAAAAAGCT ATCATATTGC TTATGAAGCA CATAAAGGTC AGTTCCGAAA	1860
	AAACGGATTA CCATACATTA TGCATCCTAT ACAAGTTGCA GGTATTTTAA CAGAAATGCG	1920
25	ATTAGACGGA CCGACGATTG TCGCAGGTTT TTTGCATGAT GTAATTGAAG ATACACCGTA	1980
	TACATTTGAA GATGTAAAAG AAATGTTCAA TGAAGAAGTT GCTCGAATTG TTGATGGTGT	2040
	GACGAAGCTT AAAAAAGTAA AATACCGCTC AAAAGAAGAA CAACAAGCTG AAAATCATCG	2100
30	CAAGTTATTT ATTGCGATTG CCAAAGATGT ACGCGTAATT TTGGTGAAAT TAGCAGACAG	2160
	ATTACATAAT ATGCGTACCT TGAAAGCCAT GCCGCGCGAA AAACAAATTA GAATTTCTCG	2220
35	AGAAACATTA GAAATTTATG CACCATTAGC ACATCGTCTT GGTATTAAATA CAATCAAATG	2280
	GGAAC TAGAA GATACGGCTC TTCGTTATAT TGATAATGTG CAATATTTTA GAATAGTCAA	2340
	TTTAATGAAG AAGAAACGTA GTGAACGTGA AGCGTATATC GAAACGGCTA TTGATAGAAT	2400
40	ACGTACTGAA ATGGACCGAA TGAATATCGA AGGCGATATA AATGGTAGAC CTAACATAT	2460
	TTACAGTATT TATCGGAAAA TGATGAAGCA GAAAAACAA TTTGATCAA TTTTGTATT	2520
	GTTGGCGATA CGTGTATTG TCAATTCTAT TAATGATTGT TATGCGATAC TTGGGTTGGT	2580
45	GCATACGTTA TGGAAACCGA TGCCAGGACG TTTTAAAGAT TATATTGCAA TGCCTAAACA	2640
	AAATTTGTAT CAGTCATTGC ATACTACAGT AGTAGGCCCA AATGGAGACC CGCTCGAAAT	2700
	CCAAATACGA ACGTTTGATA TGCACGAAAT TGCTGAGCAT GGTGTTGCAG CACACTGGGC	2760
50	TTACAAAGAA GGTAAAAAG TAAGTGAAAA AGATCAAAC TATCAAAATA AGTTAAATTG	2820
	GTTAAAGAA TTAGCTGAAG CGGATCATAC ATCGTCTGAC GCTCAAGAAT TTATGGAAAC	2880
55		

	TGAGTTGCCA	TATGGTGCTG	TGCCGATTGA	TTTTGCTTAT	GCGATTCACA	GTGAAGTAGG	3000
	TAATAAGATG	ATTGGTGCCA	AGGTGAATGG	CAAAATTGTA	CCAATTGACT	ATATTTTACA	3060
5	AACAGGCGAT	ATTGTTGAAA	TACGTACTAG	TAAACATTCA	TATGGACCAA	GTCGTGATTG	3120
	GTTGAAAATT	GTTAAATCGT	CTAGTGCCAA	AGGTAAAATT	AAAAGTTTCT	TCAAAAAACA	3180
	AGATCGTTCA	TCTAATATTG	AAAAAGGCCG	AATGATGGTT	GAAGCTGAAA	TAAAAGAGCA	3240
10	AGGATTTAGA	GTCGAAGATA	TTTTGACAGA	GAAAAATATT	CAGGTTGTTA	ATGAAAAATA	3300
	TAACTTTGCA	AATGAAGATG	ATTTATTTCG	AGCTGTAGGA	TTTGGCGGCG	TGACATCCTT	3360
15	ACAGATTGTT	AATAAATTAA	CTGAAAGACA	ACGTATTTTA	GATAAACAAAC	GTGCTTTAAA	3420
	TGAAGCACAA	GAAGTTACGA	AATCATTGCC	TATTAAAGAC	AACATCATT	CTGATAGTGG	3480
	TGTCTATGTA	GAAGGTTTAG	AAAATGTACT	TATCAAGTTG	TCAAATGTT	GTAATCCTAT	3540
20	ACCAGGTGAT	GATATTGTAG	GTTATATCAC	CAAAGGTCAC	GGTATTAAAG	TACATCGCAC	3600
	TGATTGCCCA	AATATTAAGA	ACGAAACTGA	ACGACTAATT	AATGTTGAAT	GGGTAAATC	3660
	AAAAGACGCA	ACTCAAAAAT	ATCAGGTTGA	TTTAGAGGTA	AtGCGTATGA	CCGAAATGGC	3720
25	TTGTGAATG	AAGTACTACA	AGCTGTTAGC	TCGACAGCCG	GCAATTTAAT	TAAAGTTTCA	3780
	GGACGTTTCA	ATATTGATAA	AAATGCAATA	ATAAATATTA	GTGTCATGGT	GAAAAACGTG	3840
	AATGATGTTT	ATCGTGTGGT	AGAAAAGATC	AAACAACCTG	GTGATGTTTA	TACAGTAACA	3900
30	AGAGTTTGA	ACTAGAGGTG	CAAAATATGA	AAGTAGTTGT	ACAAAGAGTT	AAAGAAGCAT	3960
	CGGTGACGAA	TGATACATTA	AATAATCAAA	TCAAAAAGG	ATATTGTTTA	TTAGTCGGTA	4020
35	TCGGTCAGAA	CTCTACAGAG	CAAGATGCAG	ATGTAATTGC	AAAGAAAATT	GCTAATGCAA	4080
	GATTATTTGA	AGATGACAAT	AATAAATTAA	ACTTTAATAT	CCAACAAATG	AATGGTGAAA	4140
	TACTATCAGT	TTCACAATTT	ACTCTCTATG	CAGATGTAAA	AAAAGGTAAC	CGTCCAGGTT	4200
40	TCTCAAATTC	TAAAAATCCT	GATCAGCGG	TAAAAATTTA	TGAGTATTTT	AATGcATGCG	4260
	CTACGAGCGT	ATGGTCTTAC	TGTGAAAACA	GGTGAATTTG	GAACACACAT	GAATGTTAGC	4320
	ATAAATAATG	ATGGTCCAGT	CACTATTATT	TATGAAAGTC	AGGACGGCAA	AATTCAATGA	4380
45	AAAAATAGA	GGCATGGTTA	TCTAAAAAGG	GTCTTAAAAA	TAAACGTACT	CTAATAGTAG	4440
	TGATTCCTT	TGTCTTATTT	ATCATCTTTT	TATTTTATT	GCTGAATAGC	AATAGTGAAG	4500
	ATAGTGGGAA	CATCACGATA	ACTGAAAATG	CTGAATTACG	TACAGGTCCA	AACGCTGCGT	4560
50	ATCCAGTCAT	ATATAAAGTT	GAAAAAGGTG	ACCATTTTAA	AAAGATTGGT	AAAGTAGGTA	4620
	AATGGATTGA	AGTTGAAGAT	ACATCCAGTA	ATGAAAAAGG	TTGGATAGCT	GGATGGCACA	4680
55							

TAGTGCTTGA TCCTGGTCAT GGAGGTAGTG ACCAGGGTGC TTCAAGCAAT ACTAAATATA 4800  
 AAAGTTTAGA AAAAGATTAT ACGTTGAAAA CAGCAAAAGA ATTGCAGCGT ACTTTAGAAA 4860  
 5 AAGAAGGCGC AACTGTTAAG ATGACAAGAA CAGACGATAC ATATGTTTCA CTAGAAAATC 4920  
 GTGATATCAA AGGCGATGCC TATTTGAGTA TACATAATGA TGC GTTAGAA TCATCTAATG 4980  
 CAAATGGAAT GACaGTTTAT TGGTATCATG ATAATCAAAG AGCTTTAGCA GATACGTTAG 5040  
 10 ACGCTACGAT TCAGAAGAAA GGTCTACTTT CTAATCGCGG TTCAAGACAA GAAAATTATC 5100  
 AAGTGTTAAG ACAAACAAAA GTTCCTGCTG TTTTATTAGA ATTAGGTTAT ATTAGTAACC 5160  
 CAACTGATGA AACGATGATT AAAGATCAAT TACATAGACA AATTTTAGAA CAAGCAATTG 5220  
 15 TTGATGGCCT TAAAATTTAT TTTTCTGCGT AGGGCTTGCA AAAATATGTG AAAGTAGTTA 5280  
 TCATTGATAT TGAATTTTAT AACTAAAACC GTTAGTATTC TTGAAATGGT AAATGAAATA 5340  
 GGTAGCAATC TAACTAAGAT TGTGTAGGAA TATAATCCAT AGACTGAAAG ATTATGCTGA 5400  
 GTAGTTTATA TACATTGAAC ACAAGAAGAG GTGCTTTATG AAAAGTAAAG CCGTTAAACG 5460  
 TACGTTaAAC GTTTTGAGTG GTTTATTAA ATGCACGCTT ATAAAAAGTA ATGATGATTA 5520  
 25 CAATTAGGCA TGT TTTTAA ACCA 5544

## (2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:  
 30 (A) LENGTH: 1067 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

AAAAGATTGC AAATATAAAT GGCATGTTTA ATATGTTAGA ACAACAAATC ATTCATAGCC 60  
 40 AAGATATGGC TCATTTTAGA AGTGAATTTT TTTACGTCAA TCATGaGCAT CGAGAAAAC 120  
 ATGAAGCACT CCTAATTTAT TACAAAAATA GTATCGACAA TCCTATTGTA GATGGTGCAT 180  
 GTTATATTTT AGCCCTACCT GAAATTTTCA ATAGTGTGTA TGT TTTTCGAA TCAGAGTTAC 240  
 45 CATT TTTTCATG GGTATATGAT GAAAATGGCA TTACCGAAAC AATGAAATCA CTTAGCATTC 300  
 CATTACAATA TTTAGTTGCA GCAGCTTTAG AAGTAACTGA TGTGAATATA TTTAAGCCTT 360  
 CAGGATT TAC AATGGGAATG AATAATTGGA ATATTGCTCA AATGCGAATC TTTTGGCAAT 420  
 50 ATACAGCAAT TATTAGAAAA GAAGCACTAT AACATTAATA ATTAATTAGC TATAAAGATG 480  
 ATTCACAACA ATCATCTTTA TAGCTTT TTTT ATGTCTAATT ATTTTTGAGG AAAATmACAA 540

55

AATTTTATGT TTTCAAAAGT AAACAATCAA AAGATGTTAG AAGATTGCTT CTATATAAGA 660  
 AAGAAAGTGT TTGTAGAAGA ACAAGGCGTC CCTGAGGAAA GTGAAATTGA TGAATATGAA 720  
 5 TCTGAATCTA TTCACCTCAT TGGATATGAT AATGGACAGC CAGTTGCCAC TGCTCGAATA 780  
 CGCCCTATTA ATGAAACAAC TGTCAAATA GAACGAGTAG CTGTGATGAA ATCACATCGT 840  
 GGACAAGGAA TGGGTAGAAT GCTTATGCAA GCTGTAGAAT CATTAGCTAA AGATGAAGGT 900  
 10 TTTTACGTAG CTACTATGAA TGCCCAATGT CATGCTATCC CATTTTATGA AAGTTTAAAC 960  
 TTAAAATGA GAGGTAATAT ATTTCTTGAG GAAGGCATCG AGCATATTGA AATGACAAAA 1020  
 AAGTTAACCT CGCTTAATTA AAAAAAGTTG TATCTATTTT AGAAACA 1067  
 15

(2) INFORMATION FOR SEQ ID NO: 112:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18613 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 20

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

AAGACGtAtG ATAACAACAA TACgTGTAGT GAAAGATTTT AATCTACATA TTACTGACAA 60  
 AGAATTCATT GTATTTGTTG GACCATCGGG ATGTGGTAAA TCAACAACAT TACGAATGGT 120  
 30 TGCTGGACTA GAGTCTATCA CATCTGGAGA TTTTATATT GATGGGGAAC GCATGAACGA 180  
 TGTGAACCA AAGAATAGAG ATATTGCGAT GGTATTTCAA AACTATGCAT TATATCCACA 240  
 TATGACTGTT TTTGAAAATA TGGCATTG GCTAAAGCTA CGTAAAGTAA ATAAAAAGA 300  
 35 GATTGAACAA AAAGTTAATG AAGCAGCTGA AATATTAGGA TTAAGTGAAT ATCTTGGTCG 360  
 TAAACCAAAA GCGTTATCTG GCGGACAGCG TCAACGTGTT GCTTTGGGCA GAGCTATTGT 420  
 TAGGGATGCG AAAGTCTTTT TAATGGATGA ACCATTATCG AATCTTGATG CGAAyTtCGA 480  
 40 GTACAAATGC GCACAGAAAT ATTGAAATTA CATAAGCGAC TTAATACTAC GACAATTTAT 540  
 GTTACACATG ATCAAACTGA AGCATTGACG ATGGCTAGTC GAATTGTTGT TTTGAAAGAT 600  
 45 GGCGACATTA TGCAAGTCGG CACACCTAGA GAAATATATG ATGCCCTAA TTGCATATTT 660  
 GTGGCGCAAT TTATCGGCTC ACCAGCAATG AATATGTTGA ATGCTACAGT TGAAATGGAC 720  
 GGATTGAAGG TAGGAACACA CCATTTTAAA TTACATAATA AAAAATTTGA AAAGTTAAAA 780  
 50 GCTGCTGGCT ACTTAGACAA GGAAATTATT TTAGGTATTC GAGCTGAAGA CATTATGAA 840  
 GAACCAATAT TTATCAAAC TTCTCCAGAG ACACAATTTG AATCTGAAGT AGTTGTATCC 900

55



## EP 0 786 519 A2

	AAATTAGATT CAAGAACTCA AGTGATGGCG AACGACAAGA TTACACTAGC ATTTGATATG	1020
	AATAAGTGTC ACTTTTTTGA TGAAAAAACA GGAAATCGTA TCGTCTAAGG GGGAGTATTC	1080
5	ATGTCTAAAA TTTTAAAATG TATCACGTTA GCCGTGGTAA TGTATTAAAT CGTAACTGCA	1140
	TGTGGCCCTA ATCGTTTCGAA AGAAGATATT GATAAAGCAT TGAATAAAGA TAATTCTAAA	1200
	GACAAGCCTA ACCAACTTAC GATGTGGGTG GATGGCGACA AGCAAATGGC GTTTTATAAA	1260
10	AAAATTACGG ATCAATATAC TAAAAAACT GGCATCAAAG TAAAGCTTGT AAATATTGGT	1320
	CAAAATGATC AACTAGAAAA TATTTTCGCTA GACGCTCCTG CAGGAAAAGG TCCAGATATC	1380
	TTTTTCTTAG CACATGATAA TACTGGAAGT GCCTATCTAC AAGGCTTAGC TGCTGAAATC	1440
15	AAATTATCAA AAGATGAGTT GAAAGGTTTC AATAAGCAAG CACTTAAAGC GATGAATTAT	1500
	GACAATAAGC AACTAGCATT GCCAGCTATC GTTGAAACAA CCGCACTTTT TTATAATAAA	1560
20	AAATTAGTGA AAAATGCACC GCAAACGTTA GAAGAAGTTG AAGCTAATGC TGCCAAACTA	1620
	ACTGATAGTA AAAAGAAACA ATACGGTATG TTATTTGATG CTAAAAATTT CTATTTTAAT	1680
	TATCCGTTTT TATTCGGCAA TGATGATTAT ATTTTCAAGA AAAATGGCAG TGAATATGAT	1740
25	ATTCATCAGC TAGGACTAAA TTCAAACAT GTCGTCAAGA ATGCTGAACG ATTACAAAAA	1800
	TGGTACGACA AAGGGTATCT TCCTAAGGCA GCAACACATG ATGTCATGAT TGGTCTTTTT	1860
	AAAGAAGGAA AAGTAGGACA ATTTGTCACT GGACCGTGA ACATTAATGA ATATCAAGAA	1920
30	ACGTTTGGTA AAGATTTAGG AGTAACAACA TTACCTACAG ATGGTGGCAA ACCTATGAAA	1980
	CCATTTCTAG GTGTACGTGG TTGGTATTTA TCTGAATATA GTAAACATAA GTATTGGGCT	2040
	AAAGATTTAA TGCTGTATAT CACTAGTAAA GATACATTAC AAAAATATAC AGATGAAATG	2100
35	AGCGAAATTA CTGGACGTGT TGACGTGAAA TCATCTAATC CAAATTTAAA AGTGTTTGAA	2160
	AAGCAAGCAC GTCATGCTGA ACCGATGCCT AATATTCCTG AAATGCGACA AGTTTGGGAA	2220
40	CCGATGGGCA ATGCAAGCAT ATTTATTTCA AATGGTAAGA ATCCTAAACA AGCGTTAGAT	2280
	GAGGCGACGA ATGATATAAC GCAAAATATT AAGATTCTTC ATCCATCACA AAATGATAAG	2340
	AAAGGAGATT AGTTATGACG AAACGTAACC CTAAATTAGC GGCATTATTA TCTGTTATAC	2400
45	CTGGTTTGGG ACAGTTTTAT AATAAAAGAC CCAATTAAAGG GACGATATTT TTTATCTTTT	2460
	TCATCAGTTT TATTTCTGTT TTTTATAGCT TTTTAAATAT TGGTTTTTGG GGATTGTTCA	2520
	CATTAGGGAC AGTACCTAAG TTAGACGATT CTCGTGTCTT ACTTGACAAA GGTATTATTT	2580
50	CTATCTTACT CGTTGCTTTC GCAATCATGC TATATATCAT TAATATTTTA GATGCATATC	2640
	GTAATGCTGA ACGATTTAAT CGCAATGAGG AAATAAAGGA TCCGAAGcGC GTATGGTGGC	2700
55		

	TG TAGTTGTA TTTCCATTAA TAyyTATGTT TGGAGTAGCA TTTACAAATT ACAATTTATA	2820
	CAACGCGCCT CCGAGACACA CATTAGAATG GGTGGTTTA GATAACTTTA AAACGTTATT	2880
5	CACAATTGGC GTTTGGCGTA AAACATTTTT CAGTGTATT ACTTGGACAT TAGTATGGAC	2940
	GCTTGTGCA ACGACACTTC AAATTGCATT AGGGCTGTTT TTGGCAATTA TTGTAAATCA	3000
	CCCTGTCGTC AAAGGTAAGA AATTTATCCG TACTGTGTTA ATCCTACCTT GGGCTGTACC	3060
10	ATCATTTGTG ACAATTTTAA TATTTGTAGC GTTATTTAAT GATGAATTTG GTGCGATAAA	3120
	TAATGATATT TTGCAACCTT TATTAGGTGT AGCACCAGCA TGGTTAAGTG ATCCGTTTTG	3180
15	GGCAAAAGTG GCATTAATCG GCATTCAAGT ATGGCTTGGA TTCCCATTTG TCTTTGCACT	3240
	GTTCACTGGA GTACTGCAAA GTATTTTCATC AGATTGGTAC GAAGCAGCAG ATATGGATGG	3300
	TGCGTCTAGT TGGCAAAAGT TTAGAAACAT CACATCCCG CATGTCATTT ACGCCACAGC	3360
20	GCCATTGTTA ATTATGCAAT ATGCAGGTAA TTTCAATAAT TTTAATCTTA TTTATCTATT	3420
	TAATAAAGGC GGTCCACCAG TGTCAGGGCA GAATGCTGGT AGTACAGATA TCTTGATATC	3480
	TTGGGTGTAT AATCTGACAT TTGAGTTTAA CAACTTCAAC ATGGGTGCAG TTGTGTCATT	3540
25	AATTATTGGA TTTATTGTTG CTATTGTCGC ATTTATTCAA TTCAGACGTA CAAGTACGTT	3600
	TAAAGATGAG GGAGGTTTAT AAGATGACAA AGAAGAAAAA CATATTAAAA GCAATCGGTA	3660
	TTTACAGTTT TATAGCGATG ATGTTTGTCA TCATTTTATA TCCACTACTG TGGACATTTG	3720
30	GCATTTCCCT TAATCCAGGT ACGAACTTGT ATGGTGCCAA AATGATACCA GACAA TGCAA	3780
	CATTTAAAAA TTATGCATTC TTA CTATTTCG ATGACAGTAG TCAATACCTG ACTTGGTATA	3840
35	AAAATACGCT TATCGTAGCA TCTGCAAATG CACTGTTTAG TGTGATATTT GTCACGTTAA	3900
	CAGCATATGC TTTTCTAGA TATCGCTTTG TTGGTCGTAA ATACGGGCTG ATTACATTTT	3960
	TGAFTTTACA AATGTTCCCT GTATTAATGG CAATGGTCGC AATCTATATT TTGCTAAATA	4020
40	CAATTGGATT ATTAGATTCT TTATTTGGAC TAACACTGGT ATATATTGGT GGATCAATAC	4080
	CGATGAATGC CTTTTAGTG AAAGGTTACT TCGATACGAT TCCAAAAGAA CTTGATGAAT	4140
	CTGCCAAAAT TGATGGTGCA GGGCATATGC GTATTTTCTT ACAAATTATG CTTCCATTAG	4200
45	CTAAGCCGAT TTAGCAGTT GTTGCTTTGT TCAATTTTAT GGGGCCATTT ATGGACTTTA	4260
	TATTACCTAA AATACTATTA AGAAGTCCTG AAAAATTCAC ATTAGCAGTT GGATTGTTCA	4320
	ACTTTATTAA TGATAAGTAT GCAAATAATT TCACAGTGTT TGCAGCAGGG GCAATTATGA	4380
50	TTGCAGTACC TATAGCAATC GTATTCTTGT TCTTGCAACG CTATTTAGTA TCAGGTTTAA	4440
	CAACAGGTGC GACAAAAGGT TAGTTTGAAA TTAGGAGTGG GGCAGAATTG ATAAAGAACC	4500
55		

EP 0 786 519 A2

	GGGTGTGGTG GTATTGCGAA TGGCAAGCAC ATGCCAAGTT TACAAAAAGT TGAAAATGTT	4620
	GAAATGATCG CATTTTGTGA CGTAGACATT TCGAAAGCAG CGAGTGC GGC AGAAGCATA C	4680
5	GGAAGT GACA ATGCAAAGGT TTATGATGAT TACAAAGCAT TGT TAAAAGA TGACACGATT	4740
	GATGTTATCC ATGTTTGTAC GCCAAATGAC TCGCATTGTG AAATTACTGT AGCAGGGTTG	4800
	CATGCTGGTA AACATGTGAT GTGTGAAAAA CCAATGGCTA AAACGACAGC AGAAGCTCAA	4860
10	AAAATGATAG ATACAGCTAA ATCAACAGGT AAAAAATTAA CAATAGGTTA TCAAAATCGT	4920
	TTCCGAGCAG ATAGTCAATT TTTACATCAA GCAGCGCAAC GTGGCGACTT AGGAGACATT	4980
15	TACTTCGGAA AGGCACATGC CATTCTCGT CGAGCAGTAC CAACATGGGG TGTCTTTCTA	5040
	GACGAAGAAG CTCAAGGTGG AGGACCATTA ATCGATATCG GTACACACGC TTTAGATTTA	5100
	ACGTTATGGA TGATGGATAA TTATGAACCA GAATCAGTGA TGGGTTCAAC ATTCCATAAA	5160
20	TTAAATAAAC AGCATCATGC GGCAAACGCT TGGGGTTCAT GGAATCCAGA TGAATTTACA	5220
	GTTGAAGATT CTGCGTTTGG ATTTATTAAA ATGAAGAATG GAGCGACGAT CATTTTAGAA	5280
	TCCGCTTGGG CGATTAATTC TTTAGAAGTG GATGAGGCAA AATGTTCAAT ATCAGGAACT	5340
25	AAAGCAGGTG CTGATATGAA AGATGGTCTA CGTATTCATG GTGAAGACAT GGGTACACTT	5400
	TATACCAAAC ACGTTGAATT GGAAAACAAA GCGCTCGACT TTTATGAAGG TAATGAAGTG	5460
	GATGAAGCTG AAGAAGAAGC AAAAGCTTGG ATTGATGCAG TTGTAAATGA TACTGAACCA	5520
30	GTTGTGAAAC CGGAACAAGC AATGGTAGTT AAAAAATTC TTGAAGCGAT TTATCAGTCT	5580
	GCAAAATCAG GCAAAGCAAT TTACTTTGAA TAACATCATA CGGTAAGGAG GCACATCATG	5640
35	ACAAAATTAA AAGTTGGTGT GATAGGTGTT GGTGGTATTG CACAAGACCG TCATATTCCA	5700
	GCATTGCTGA AACTCAAAGA CACAGTCTCA TTAGTTGCAG TACAAGATAT TAATACAGTG	5760
	CAGATGATTG ATGTTGCGAA GCGCTTTAAT ATACCTCATG CAGTTGAGAC ACCTAGCGAG	5820
40	CTGTTTAAAC TTGTTGATGC GGTGGTCATT TGTACACCTA ATAAATTCCA TGCTGATCTT	5880
	TCTATAGAAG CATTGAACCA TGGTGTCCAT GTATTGTGTG AAAAGCCAAT GGCGATGACG	5940
	ACGGAAGAGT GTGATCGCAT GATTGAAGCG GCTAATAAAA ATCACAAATT ATTAAGTGTG	6000
45	GCATATCATT ATCGTCACAC AGATGTGGCA ATTACTGCTA AAAAAGCAAT TGAATCAGGT	6060
	GTGGTTGGTA AACCTTTAGT AGCACGTGTA CAAGCGATGC GTAGGCGTAA AGTGCCTGGC	6120
	TGGGGTGT TTACCAATAA AGCGTTGCAA GGTGGCGGTA GTTTAATCGA TTATGGTTGC	6180
50	CACTTGTTAG ACTTATCTTT GTGGCTACTA GGTAAAGATA TGGTGCCGCA TGAAGTGCTA	6240
	GGAAAAACAT ATAATCAATT GAGCAAACAA CCGAATCAAA TTAATGATTG GGGAACATTT	6300
55		

	GCAAGCATGC	AGTTTGAATG	TTCGTGGTCT	GCAAATATCA	AAGAAGATAA	GGTTCACGTT	6420
	AGTTTATCAG	GAGAAGATGG	CGGTATCAAT	TTATTTCCAT	TTGAAATATA	TGAGCCCCGC	6480
5	TTTGGAAC TA	TTTTTGAAAG	CAAAGCTAAT	GTTGAGCATA	ACGAAGACAT	TGCTGGTGAG	6540
	AGACAGGCGC	GTAAC TTTGT	CAATGCGTGT	TTAGGGATAG	AAGAGATTGT	GGTGAAACCG	6600
	GAAGAAGCAC	GCAATGTAAA	TGCCCTTATA	GAAGCGATTT	ATCGTAGCGA	TCTTGATAAC	6660
10	AAGAGCATAC	AAC TTTAATG	ATTATCATAT	ATGATACAAA	ATTCTCAATA	TAAAAAGAAG	6720
	GAGTGCTTTT	CAATGAAAAT	AGGTGTATTT	TCAGTATTAT	TTTACGATAA	AAATTTTGAA	6780
15	GATATGTTAG	ATTATGTCTC	AGAATCTGGA	TTGGATATGA	TTGAAGTTGG	AACAGGTGGT	6840
	AACCCAGGAG	ATAAATTTTG	TAAGTTAGAT	GAGTTGTTAG	AAAATGAAGA	CAAGCGCCAA	6900
	GCATTTATGA	AGTCAATCAC	AGACAGAGGC	TTACAAATAA	GTGGTTTCAG	TTGTCATAAC	6960
20	AATCCAATTT	CTCCAGATCC	GATAGAAGCG	AAAGAAGCCG	ATGAAACGTT	ACGTAAAACA	7020
	ATCCGTTTAG	CAAATCTATT	AGACGTGCCA	GTTGTTAATA	CATTTTCTGG	CATTGCAGGA	7080
	TCAGATGATA	CCGTAAAAA	GCCTAATTGG	CCTGTTACAC	CTTGCCCAAC	AGCCTACTCT	7140
25	GAAATTTATG	ATTATCAGTG	GAATGAAAAG	TTGATACCAT	ATTGGCAAGA	TTTAGCTGAG	7200
	TTTGCAAAAAG	AGCAAGATGT	AAAAATTGCC	ATAGAGTTGC	ATGCAGGATT	TTTAGTGCAT	7260
	ACACCATATA	CAATGTTGAA	GTTACGTGAG	GCTACAAATG	AATATATCGG	TGCTAACTTA	7320
30	GATCCTAGTC	ATCTATGGTG	GCAAGGTATT	GACCCAATTG	CTGCGATTCT	CATATTAGGC	7380
	CAAGCAAATG	CAATTCATCA	CTTCCATGCT	AAAGATACGT	ATATTAATCA	AGAAAATGTA	7440
35	AATATGTATG	GTCTAACTGA	TATGCAACCA	TATGGTAACG	TTGCGACAAG	AGCATGGACA	7500
	TTCCGTACAG	TTGGTTATGG	ACATAGTCCA	TATGTATGGG	CAGATATCAT	AAGTCAACTT	7560
	ATTATTAATG	GATATGATTA	TGTATTAAGT	ATTGAACATG	AAGATCCTAT	TATGTCAGTA	7620
40	GAAGAAGGTT	TCCAAAAAGC	TTGTCAAAC T	TTGAAATCTG	TTAATATTTA	CGACAAGCCA	7680
	GCAGACATGT	GGTGGGCATA	ATACGAACTC	GAGGTTAGTC	TGAAGTTTGT	CTGAAGTAAG	7740
	ACTGGTGGCA	GTGTTGAATA	AATGCATATG	TCGCCAAGCC	ATTGCCAAAA	ATTTACACCC	7800
45	TTAAATCAAG	TCATTGTTTG	TAAAGAAGGT	GTA CTTTATA	TAAGTATATA	GCGATGGTCA	7860
	TACCCATTCA	CAGTAACAAT	CCTCACCATT	GAAAAGAGTA	TATAACCTTT	TCAATAGTGA	7920
	GGTATATGAT	AATAAAAAAA	GCCTGTTGTC	ACAATGGTCA	TAGACACGAC	ATACTTTAAA	7980
50	GGTTTCTGAA	TATAATATTT	CAGAATGCAC	TTTAAAGATG	GACGTCGATG	TAGACTAAAG	8040
	TGATGACAGG	CTTTCATCTT	TTTAAATATT	CATTAATTTT	TCTTCTTGTT	TAATACGTAC	8100
55							

	TAATACACCG ATTAATTCAG GAATGATGTT TAAGAAGTAA TTTGGGTGTT TTGTAATTTT	8220
	ATATAATCCA GATTTAATAA TAGGATGGTT AGGTAAATG AATAATTTTA ATGTCCAAAT	8280
5	ACCACCTAAA GTTTTAATAA CCATAAATAA CATGATATAA GCAAAGATTA ATATAACTAA	8340
	GCCAATACCA TTTGCAAAGC TAAATGTATC TTTATTAATA AATGCCTCTA CACCAGCCAA	8400
	TACATAAATT AAAACGTGTG TTATTGCTAA AAACCTCGAA TTTTAAACGC CATATTC AAC	8460
10	TGCACCGTCT GCTTTTAATT GTTTTGAGTG ATTAATAGAT ATCTTTAAGC TGACAAGTCT	8520
	GATACAGAAA AAGATAAGTA ATATAGATAG AATCATGATG TCCTCCGTCA TTATGTCATA	8580
15	TGTATAAGCG TTGATTTTGA CAACATAAAG TATTTTATAG ATAAAGCTTG TCAAATACTA	8640
	TTAACTATTT ATTAATTTTA GTACATAAAT ATGTTTCTAA GTATGTGTTT ATGTT CAGTA	8700
	TTTTGGATAA TTTAATAATT TTAAGGATAT TAAGCGCTTA CACCGACGTG ATATATTTGG	8760
20	CTTAACGAAA ATGATTGAGG TGACAGAGAT GAACTTTTTT GATATCCATA AGATTCCGAA	8820
	CAAAGGCATT CCATTATCGG TACAACGTAA ATTATGGCTT AGAAACTTCA TGCAAGCTTT	8880
	CTTCGTAGTG TTCTTTGTTT ATATGGCTAT GTATTTAATT CGAAACA ACT TTAAGGCGGC	8940
25	ACAACCGTTT TTAAGAGAGG AAATTGGATT ATCTACATTA GAACTTGGTT ATATCGGATT	9000
	AGCATTTAGT ATCACGTACG GTTTAGGAAA AACATTACTT GGATATTTTG TCGATGGACG	9060
	TAACACAAAA CGTATTATCT CGTTCTTACT TATCTTATCT GCGATTACAG TTTTAATTAT	9120
30	GGGATTTGTT TTAAGTTACT TTGGTTCTGT AATGGGATTA TTAATTGTAC TTTGGGGACT	9180
	TAACGGGGTG TTCCAATCAG TTGGTGGACC TGCAAGTTAT TCAACGATTT CAAGATGGGC	9240
35	GCCAAGAACG AAACGTGGCC GATACTTAGG ATTCTGGAAT ACATCACATA ATATCGGTGG	9300
	TGCCATAGCA GGTGGTGTG CACTTTGGGG TGCTAATGTA TTCTTCCATG GAAATGTTAT	9360
	AGGGATGTTC ATTTTCCCAT CGGTGATTGC ATTACTTATT GGTATCGCAA CATTATTTAT	9420
40	CGGAAAAGAT GATCCGGAAG AATTAGGATG GAATCGTGCT GAAGAAATTT GGGAAGAGCC	9480
	GGTCGATAAA GAAAATATTG ATTCTCAAGG TATGACGAAA TGGGAGATCT TTAAAAAATA	9540
	TATCCTGGGA AATCCTGTTA TATGGATTCT ATGTGTTTCA AACGTCTTTG TATACATTGT	9600
45	ACGAATCGGT ATTGATAACT GGGCACCGTT ATATGTGTCA GAGCATTTAC ACTTTAGTAA	9660
	AGGCGATGCA GTTAATACGA TATTCTACTT TGAAATTGGT GCATTAGTTG CAAGTTTATT	9720
	ATGGGGCTAC GTATCAGACT TATTAAAAGG TCGTCGTGCA ATTGTAGCTA TTGGCTGTAT	9780
50	GTTTATGATT ACATTTGTTG TCTTATTCTA CACAAATGCT ACAAGTGTCA TGATGGTTAA	9840
	CATTTCAITG TTTGCATTAG GTGCGTTAAT CTTTGGTCCG CAATTATTAA TTGGTGTATC	9900
55		

	CGCGTATCTA TTCGGTGACT CAATGGCGAA AGTTGGTTTG GCGGCTATTG CTGATCCAAC	10020
	ACGTAACGGT TTAAACATCT TTGGATATAC ATTAAGTGGA TGGACAGATG TTTTCATCGT	10080
5	CTTCTATGTT GCATTATTCC TAGGCATGAT TCTATTAGGA ATCGTTGCTT TCTATGAAGA	10140
	AAAGAAAATT AGAAGTTTAA AAATTTAATA TAAATCGGAT TAAAAGTATC GCCAATCTAT	10200
	TGCAATATAG TTGGCAATCC TGCCCCGACG GCATGTGCGT GAAGAGATGA AAGATACTGC	10260
10	TTCTACCCTT GCAAATATAT CATCTCTATG TCTCGGGGCA GATCATAATT CCCTGTTATG	10320
	AAGTATCCTT ATTTGCCCGA CTTAGGGTGA CTCAATGAAT TTAATCCTTA CAATAAAGAC	10380
15	ATATAGCGGT GTCAATATTG TAGGGAGTAT TGTTTTATAT TTAAACTCTC TAAAAAGCGG	10440
	ACTGAAAGAA AAGTGAAAAC TTCTCTATCA GTCCGCTTTT TCATAGAACA AAATGGAGGC	10500
	GCCATAATCA TTAGTTATGT GCTAATCTAT TTTGCTTGCT TACAATAATC ACTTGGCGAC	10560
20	ATTTGTAAAT ATTTTTTAAA ATGATAGCTA AACATTTTAT ACTCTGAAAA GCCTACTTTG	10620
	TCTGCAATTT CATAGTGTTT GTAATGTCGA TCTAACAATT GCAGAGATTG TAAAAACGA	10680
	TAGCGATTTA AATAATCGAC AATTGTAATA CCAACATGAT CTTTAAATGT TCGCATCGCA	10740
25	TACGATTCAC TAACATCGAT ATGTTGAATT AAATCTGAAA CAGTCACTTT CGTTTGATAA	10800
	GATTGCTTAA TTTGATCCAC AATCTGGTTT ACATAATAAT CATCGTATTC TACTTTTAAT	10860
	AGTGGTTGGA AGGCATCATG ACAAGATGCT AAGCTACGGC CGTTCTGTGA TTGTTGCTCT	10920
30	AATAAGGTAC GGACAAGTCT TCCTAAAATA ACTTCTAATT GTGCATGGTC TACTGGTTTT	10980
	AATAAATAAT CAAGAACATG ATGTTGAATG CCGGCTTTCA TATATTCAAA GTCATCGTAA	11040
35	CTCGATAATA TGATGACATT ACAATCTAGA TCGCCAATAT CATTGAGTAA ATCGACGCCA	11100
	TTTTTACGTG GCATACGAAT ATCAGTAATT ACTAATTCTG GCTGATGTTG TTGAATTAGT	11160
	GATAATGCTT CAACACCATC TTTAGCAGTG TATATTGTAT TGAAATGATA GTCTCCCCAA	11220
40	GGAAATGATTT GCTTTAATCC TTCTCGAATA ATTCGTTTCAT CATCACAAAT AACTACCTTA	11280
	AACATCTACA TTCCCCCTTG AAAGTGGTAT TTTATAACAA ATTAACGTAC CTTGATTACG	11340
	CTTTGAAAAA ATATGGAGTC GTGCATGTGA ACCATATTGA ATCATTGCTT TATTGTGTAA	11400
45	ATGATTTAAT CCCAAATGCT TAGTATCAAA TACATCATTA TTAAGAGATT GCGGTACATA	11460
	TTGCAGGCGA GATGACGACA TCCCAGATACC ATTGTCGCAA ACTAAAACAT GTAAATTCTG	11520
	ACGTGCCAAT GTCAGGCGTA TAGTAATGTC CAATGACTCA GTATCTCTAC CATGTTTAAT	11580
50	AGCATTTTCT ATGAGTGGCT GAAGCATCAT TTTACCAATT GTCTGGTGAC GCGCTTCTTC	11640
	AGAACTTTCA ATATGGAGCT TAATCATGTC ATCAAAACGG aTGTTTTGTA TTGCAACATA	11700

55

	GTAACGTAAC ATTTGCGATA ATTGTTGGAC CACAGTTtGT GCTAATTTTCG GAGATAACGT	11820
	AATTAAATAT TGTATTGTTT GCATCGTATT GAATAGGAAA TGAGGCTGGA ATTGGCGTTC	11880
5	TATTTCCCTT AACTGAATAT CACGCAAGCG ACGTTCTGTA TGCTCGATAG AATGGATCAG	11940
	TTGCTCATTT GATTCAAATA AATCGTAAAT ATAATTATTA ATTTCTTCTA GTTCACTGTT	12000
	GTTTTTTAAA GCGGTATATG TACCTAGATG ACGATTTTTG GCATAGTAAA TTTTTTGAAT	12060
10	AATCGTTTCG ATATCTTTTG TTTGTCGTTT AGCCATATTA TCTGCGCTAA TGAAACCAAA	12120
	TATTACTAGT AAAACAAGAA CTACGGCCAT AACAAITTAAC AACGTGATAC CATCTTCAAT	12180
15	GTTTTCATGT ATATCTTTAT AAATAATGAG ACGATGGTCA GCATGGTTTA ATTTTACAGA	12240
	TTCATTGATA AATCCGAATT GTTGTGGTtT ATACTTTTCA CCTATAGTAA AACGGTCATC	12300
	GTTGGCGTAT AAAATATTGT CATATTGATC AmCGATAAGT GCGAATTGTC GGTtATCTTT	12360
20	CtTAATTTCA CTAAACGTG GGGTGTtAGC CATATAAAIt TTaAGCATAT ATGTACTATT	12420
	TTTGAATTTA AGCTGATGCG TTGAAAATAA ATACATATTT TTAGTGTtTA AATGTTGATA	12480
	ATTATTGGTT ATAAACTGAT TTGGTCCAGA TAATTCATAA TAAAGTGTG CGGGCTGTTG	12540
25	GkGTATTAAT TTTAATAATT CACGTTTTGT AGCGGTCACA TCATGATGAT TTGyTAAATC	12600
	GAGCTCTTGA AACGAATTAT TATGCTGTGT AATAAATGTC TGAATCTGCT TTTCAGTATG	12660
	ATGTAAAGAT GACTGACTTT CATCAACATG TTGATGAATC GTACGATGCT CAATCCAAAT	12720
30	ATAGATGGCA TAGAAGCTTA CTAGTCCAAT AATAATGACT AAAAATACTG GAAAAATAGT	12780
	AGACnCAAAT AACGATCGTC TTAATTGATG TCTATAAGGT TTGTATGCCn TCATTGAATC	12840
	ATCTCCAAAA ATTTATGATG TGGAATATCC GGTAATTTAG ATTTCGGTAT TAAAGGTATG	12900
35	TTCTTAAGAT TTTGATAGA CTGATCGCTT TGTTCACTAA CATCCTTTTCG AATTGACTTG	12960
	GCATtGAACt CTGCAACTAA TCGTtGTTGT ACTGAGCGGC TTGTTAAATA TTGCACTAAC	13020
40	TTTTTACGCT TAGGATGAGG GTGTGCATTT TTAATAAAG CAATrCCATC AACATTTAAC	13080
	ATTGTTCCCTT CAATTGGATA AACGATTGAT ACAGGATAAC CTTTGTTTTT CCATGTGCGT	13140
	GCATCTTGTT CGTAGCTTAG ACCTGCGTAA TATTTACCTT TTGCAACATC TTCAATGACT	13200
45	TTAGACGTCT TTGACAGTTG CATCGCATGG TTTTGAATT GATGCACATC ACTTACTCGA	13260
	TGATGCATGC TATAAATAGC ACGCATATGT TGATAGCCTG TCGTTGTTGT ATTTGGATTt	13320
	GAGTACGCAA TTTTACCTTT AAGTATAGGT TGTAATAAAT CTTGATAACC TCGAATCTTA	13380
50	ATATCTCCTT GTAAATCTGA ATTCACTACT ATAAGTGTG GCATTAATAG AAAACTAGTA	13440
	ACATATTTAT TGTTGAGCG ATAATCCTCT AATTGCTGTG TTACAGATGT ATCTTGATAG	13500

55

	CCACGCTCCG	AAAAATCTTC	GTTATGCAAG	TTTGAAAGCA	GTACTTGAGT	AGATCCGTGT	13620
	TTAATTTCAA	TTTTGACATG	CTCTTGTTT	TCAAATTCAT	TTAAAATTGG	ACGAATCAAG	13680
5	TTTGATTGAT	ACGGAGAATA	AACGTGTAAT	ACATTTTAT	CGGATTCAGA	GTGACGCGTA	13740
	TTAGCGCATG	CTGaTAAAAA	AATGAGAAAT	AATAGCAAGA	TATAAATTTT	TGATTTTCATG	13800
10	ATATCCCATC	AATTCTATGT	ATATTTTAAT	ACAATAATTT	TAGCAATAAA	TGACGCATAA	13860
	GTAATGTAA	ATATTTAGAA	ATGTTTATAG	ATGACTTGTT	AAGACGTTGC	AAATGTTGTG	13920
	ATAGCACAAA	ATTTTTGTTT	GTCAAGACGA	TTTACCGAGG	CTGTAAAATC	AAACTGTTAT	13980
15	ATTTTATTG	TAGCTGTTAT	ATAAAAATCG	GCAAGATATT	GAACGTTCA	AAAGTGAATT	14040
	TTTACGTCAA	TAAAAGTATT	TAATCCAGTC	TCTTCATATA	TAAAAGTAAA	TCTTTCTAAG	14100
	TGTTGATTTA	ACGCTTATCA	ACAATCATT	TTTATAAACA	AATATATACT	CCTAAATTAA	14160
20	CTTTTAAAGC	AATGAAAATA	GTGAACATTA	TAACTGTTGT	GTAACAGAAT	GCAATTAGCA	14220
	TATTACTGTT	ACACAAATTA	GTACAGTTTC	TATGTTTGA	CATACATTG	ATGAAAATTG	14280
	TACATAATTT	ATGTGAAAAA	AATCACAACA	AACATGCTAC	AATGACTATG	AAAACGTTAA	14340
25	CATAGCATT	CAAATTCACA	ACATTATACA	GATGGAGGCG	TTTAGTATGT	TAGAAACAAA	14400
	TaAAAAATCAT	GCAACAGCTT	GGCAAGGATT	TAAAAATGGA	AGATGGAACA	GACACGTAGA	14460
	TGTAAGAGAG	TTATCCAAT	TAAACTACAC	TCTTTATGAA	GGTAATGATT	CATTTTITAGC	14520
30	AGGACCAACA	GAAGCAACTT	CTAAACTTTG	GGAACAAGTA	ATGCAGTTAT	CGAAAGAAGA	14580
	ACGTGAACGT	GGCGGCATGT	GGGATATGGA	CACGAAAGTA	GCTTCAACAA	TCACATCTCA	14640
35	TGATGCTGGT	TATTTAGACA	AAGATTTAGA	AACAATTGTA	GGTGACAAA	CTGAAAAGCC	14700
	ATTCAAACGT	TCAATGCAAC	CATTCGGTGG	TATTCGTATG	GCGAAAgcAG	CTTGTGAAGC	14760
	TTAÇGGTTAC	GAATTAGACG	AAGAACTGA	AAAAATCTTT	ACAGATTATC	GTAAAACACA	14820
40	TAACCAAGGT	GTATTCGATG	CATATTCTAG	AGAAATGTTG	AACTGCCGTA	AAGCAGGTGT	14880
	AATCACTGGT	TTACCTGATG	CATACGGACG	TGGACGTATT	ATCGGTGACT	ATCGTCGTGT	14940
	AGCTTTATAT	GGTGTAGATT	TCTTAATGGA	AGAAAAATG	CACGACTTCA	ACACGATGTC	15000
45	TACAGAAATG	TCAGAAGATG	TAATTCGTTT	ACGTGaAGAA	TTATCAGAAC	AATATCGTGC	15060
	ATTAAGAGAA	TTAAAGAAC	TTGGACAAAA	ATATGGTTTC	GATTTAAGCC	GTCCAGCAGA	15120
	AAACTTCAAA	GAAGCAGTTC	AATGGTTATA	CTTAGCATAC	CTTGCTGCAA	TTAAAGAACA	15180
50	AAACGGTGCA	GCAATGAGTT	TAGGTCGTAC	ATCAACATTC	TTAGATATCT	ATGCTGAACG	15240
	TGACCTTAAA	GCAGGCGTTA	TTACTGAAAG	CGAAGTTCAA	GAAATTATTG	ACCACITCAT	15300
55							



EP 0 786 519 A2

	AGACCCAACT TGGGTAAGT AATCTATCGG TGGTGTAGGT ATTGACGGAC GTCCACTTGT	15420
	TACGAAAAAC TCATTCCGTT TCTTACACTC ATTAGATAAC TTAGGTCCAG CTCCAGAACC	15480
5	AAACTTAACA GTATTATGGT CAGTACGTTT ACCTGACAAC TTCAAAACAT ACTGTGCAAA	15540
	AATGAGTATT AAAACAAGTT CTATCCAATA TGAAAATGAT GACATTATGC GTGAAAGCTA	15600
	TGGCGATGAC TATGGTATCG CATGTTGTGT ATCAGCGATG ACAATTGGTA AACAAATGCA	15660
10	ATTCTTCGGT GCACGTGCGA ACTTAGCTAA AACATTACTT TACGCTATCA ATGGTGGTAA	15720
	AGATGAAAAA TCTGGTGCAC AAGTTGGTCC AAACCTCGAA GGTATTAAAC GCGAAGTATT	15780
15	AGAATATGAC GAAGTATTCA AGAAATTTGA TCAAATGATG GATTGGCTAG CAGGTGTTTA	15840
	CATTAAGTCA TTAAATGTTA TTCACTACAT GCACGATAAA TACAGCTATG AACGTATTGA	15900
	AATGGCATTG CATGATACAG AAATTGTACG TACAATGGCA ACAGGTATCG CTGGTTTATC	15960
20	AGTAGCAGCT GACTCATTAT CTGCAATTAA ATATGCACAA GTTAAACCAA TTCGTAACGA	16020
	AGAAGTCTT GTAGTAGACT TTGAAATCGA AGGCGACTTC CCTAAATACG GTAACAATGA	16080
	CGACCGTGTA GATGATATTG CAGTTGATTT AGTAGAACGC TTCATGACTA AATTACGTAG	16140
25	TCATAAAACA TATCGTGATT CAGAACATAC AATGAGTGTA TTAACAATTA CTTCAAACGT	16200
	TGTATACGGT AAGAAAAGT GTAACACACC AGACGGACGT AAAGCTGGCG AACCATTTGC	16260
	TCCAGGTGCA AACCCTATGC ATGGCCGTGA CCAAAAAGGT GCATTATCTT CATTAGTTC	16320
30	TGTAGCTAAG ATCCCTTACG ATTGCTGTAA AGATGGTATT TCAAATACAT TCAGTATCGT	16380
	ACCAAAATCA TTAGGTAAAG AACCAGAAGA TCAAACCGT AACTTAACTA GTATGTTAGA	16440
	TGGTTACGCA ATGCAATGTG GTCACCACTT AAATATTAAC GTATTTAACC GTGAAACATT	16500
35	AATAGATGCA ATGGAACATC CAGAAGAATA TCCACAGTTA ACAATCCGTG TATCTGGTTA	16560
	CGCTGTTAAC TTCATTAAAT TAACACGTGA ACAACAATTA GATGTAATTT CTCGTACATT	16620
40	CCATGAAAGT ATGTAACAAA ATTTAAGGTG GGAGCACTAT GCTTAAGGGA CACTTACATT	16680
	CTGTCGAAAG TTTAGGTACT GTCGATGGAC CGGGATTAAAG ATATATATTA TTTACACAAG	16740
	GATGCTTACT TAGATGCTTG TATTGCCACA ATCCAGATAC TTGGAAAATT AGTGAGCCAT	16800
45	CAAGAGAAGT CACAGTTGAT GAAATGGTGA ATGAAATATT ACCATACAAA CCATACTTTG	16860
	ATGCATCGGG TGGCGGTGTA ACAGTCAGTG GTGGCGAACC ATTGTTACAA ATGCCATTCT	16920
	TAGAAAAATT ATTTGCAGAA TTAAAAGAAA ATGGTGTGCA CACTTGCTTA GACACATCGG	16980
50	CTGGATGTGC TAATGATACA AAAGCATTTT AAAGGCATTT TGAAGAATTA CAAAAACATA	17040
	CAGACTTGAT ATTATTAGAT ATAAAACATA TTGATAATGA CAAACATATT AGATTGACAG	17100
55		

TATGGATTTCG ACATGTCCTT GTGCCTGGTT ATTCTGATGA TAAAGACGAT TTAATTAAAC 17220  
 TAGGGGAATT TATTAATTCT CTTGATAACG TCGAAAAGTT TGAAATTCTG CCATATCATC 17280  
 5 AGTTAGGTGT TCATAAGTGG AAAACATTGG GCATTGCATA TGAATTAGAA GATGTCGAAG 17340  
 CGCCCGATGA TGAAGCTGTT AAAGCAGCCT ACCGTTATGT TAACTTCAA GGGAAAATTC 17400  
 CCGTTGAATT ATAAATACAA TTCAGACCGA AAAGAAAGCA TATGCAACTT CAAGAGTGAA 17460  
 10 GGGGCATATG CTTCTTTTTC AATTGAGTAT TGAGTATTAG CAAGACGTAG TAAGTATATG 17520  
 AGACAACTTC TACAATGGTT GAAGGAAGAC GTTTTGTAA GTAGCTATGC TGATAAAGAA 17580  
 TGTGATGTCT TGTAAAGGT GGGGTTCCTA TATCATCATT TAGCTGATGT TGAATGGGTT 17640  
 15 ATTATTTGCT ACTTGCATAT GAATATGAGT CTTTCAAAT TTTTATTGAC CCTGAGTAAT 17700  
 GAAAAATATT AAGATGAAAC TTAATATTAA AgCAATGCGG AGCGTGATTA TGAAGAGAAT 17760  
 20 TAGTAAAGAT ATATGGGCAG TATTTAAATT ACTGTATCaA AATAAAGGGC GTTTTAGCAT 17820  
 TAATGCCTTA CTATTGCAGT TAATCATGAT TTTTATTAGT AGTACATACT TAATTTTACT 17880  
 ATTTAATATG ATGTTAAAAG TAGCTGGcAA AGCCAACTTA CGATTAAcAA TTGGACGGAA 17940  
 25 ATCGTTAGTC ATCCCGCCAG TGTGATACTT CTTATTATAT TCATATTAAG TGTGCCTTT 18000  
 CTGATTTATG TAGAGTTTTT ATTGTTAGTT TATATGGTTT ATGCCGGCTT TGATCGACAG 18060  
 ATTATTACAT TTAAATCCAT TTTTAAAAAT GCCTTTGTAA ATGTGCGTAA ACTCATAGGT 18120  
 30 GTACCAGTTA TTTCTTTTGT CATTTATTTA ATGTTAATGA TACCCATTGC CAACCTAGGA 18180  
 CTAAGTTCAG TATTAACAAA AAATATTTAC ATACCTAAAT TTTTAACGGA AGAACTTATG 18240  
 AAAACGACGA AAGGTATAAT CATTTACGGT ACCTTTATGA TTGCTGTATT TATATTAAAT 18300  
 35 TTTAAATTAA TATTTACTCT ACCGTTAACG ATTTTAAACC GCCAGTCGTT ATTTAAAAAT 18360  
 ATGAGACTAA GTTGGCAAAT TACGAAGCGA AATAAGTTTC GGCTTGTTAT AGAAATAGTT 18420  
 ATATTAGAAC TCATCATTGG TCGATTTTA ACATTAAATTA TTTCAGGAGC AACATATCTT 18480  
 40 GCTATTTGTG TAGATGAAGA AGGAGATAAG TTTTtAGTCT CATCAATTTT ATTTGTTGTA 18540  
 TTGAAAAGCG CATTGTTCTT CTATTATkTA TTtACGAAAT TATCATTAAAT CAGTGTGTTA 18600  
 45 GTACTGCACT TAA 18613

(2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1214 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

EP 0 786 519 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

	AAAGTTTAA AAGGGGTGAG ATACTTGGCG AATAATCCAT TCCAGCTTG CGTTTAAAAG	60
5	GAATTATACT TGCCATTGTC GGTGCTTGTT TATGGGGATT AGGTGGTACT GTTCTGATT	120
	TCTTGTTCAA ATATAAGAAT ATTAATGTCG ATTGGTACGT CACTGCTCGA CTTGTAGTCA	180
	GTGGTGT TTT CTTACTTATT ATGTACAAA TGATGCAACC CAAACGTTCA ATATTAGCG	240
10	TATTCCAAGA TCGACGTATG TTAGGCAAAT TACTTATCTT CAGTATACTG GGCATGTTAG	300
	TAGTACAATA TGCTTATATG GCATCTATTA ATACAGGTAA TGCTGCGATT GCAACATTAC	360
	TACAATACAT TGCGCCAGTT TATATTATTA TTTGGTTTGT CATAAGAGGC GTTGCAAAAC	420
15	TAACATTATT TGATGTGCTT GCTATTATCA TGACACTATT AGGAACATTT TTATTATTAA	480
	CAAATGGTTC ATTTTCTAAT TTAGTCGTCA ATCCTGCAAG TTTATTCTGG GGTATTTTAG	540
	CTGGTGTAGC ACTCGCTTTT TACACAATTT ATCCTTCAGA CCTACTTAAC CGCTTCGGTT	600
20	CGATTCTAAT TGTCGGGTGG GCAATGCTTA TTTCTGGTGT TGCGATGAAT TTACGCCATC	660
	CAATTTGGCA CATTGATATC ACTAAATGGG ACATATCAAT TATATTATTT TTAATCTTTG	720
	GTATTATCGG TGGTACCGCA CTCGCATTTT ATTTCTTTAT CGACAGTTTA CAATACATAT	780
25	CAGCGAAAGA AACACATTA TTCGGAAGTGT TTGAACCTGT CGTAGCCGTT ATCGCAAGCA	840
	GTCTATGGTT ACATGTGGCA TTCAAACCAT TTCAAATCGT AGGCATCATT CTTATTATGA	900
30	TTTTAATTTT ATTACTATCA CTTAAAAGAC AACCTGAAAC ATTAGATGAA TAAGAAAAC	960
	CTGATAATCA CTTTAGCAAG TAACTATTAT TTAACAACGT AGTTACCTTA TAGGTGATAT	1020
	CAGAGTTTTT TATTTTAGTT AATAATATTT TTCACTTGGT ATAAAAAGC GTCGTCGCTC	1080
35	TGGTAATCGG AAATACTGGA ATAAAATATG GAATTGGGTA ATAATCCCAG GTAnTAAAAG	1140
	TCCATGTTC GATAnCCTnT CCGCAnCTCC AACCAAATTT GCCGATAAGG TTCCAAAAGG	1200
40	CATCCTGGGG GTAC	1214

(2) INFORMATION FOR SEQ ID NO: 114:

(i) SEQUENCE CHARACTERISTICS:

	(A) LENGTH: 9458 base pairs
45	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

	ATTTTGGTTT CATTACAGAT GGGGTnATAC AGCAAACACA nCTAAAATAA CTATCAATAG	60
--	---	----

55

	CTTAGACAAT AAAAAATATG CCACTACAAT CGCTAATATT ACGATTAAAA AAGAAGCGTT	180
	AACGATTACT TTCATCGTTG TTCTATCTCT GAACATCATA TTAAAGACAA CTAGACTAAT	240
5	TGATAATGAA ACAGCAAAAA AAGTAATAGC TAACACTAAT TTCATCATAA ATAGACAGAC	300
	TAAACCTATG ACTAATAATG TATTAGAAAT TACAGCTGAC GTTTTAAACA TTCTCGaATT	360
	AATATGCACT CACCCTTTTT ATTTAAATAA CTTACATAAT CATAATAATA CATGATGTTT	420
10	CATAGGCCTG TCGATGATTG ATTCACAATA GCACGTGATT TTTTGTGTTT TCAATATTAT	480
	TCATTTATTC CATCAAAAAC ACCCTTTTTA ATTTTACAA AAATTAAAAA AAGTGCTCCT	540
15	ACACTGCTTG CATGTAGAAA CACTTTTTCA TTGTAATGTT ATTCTTCTCG AGACATACCT	600
	TTTAGCATAT TAAGCATGTA TGTTAAACTA CGGTTTCATG CGTCATCTTT CAATACGCCC	660
	AATAGACTTC TTATAGTTGT CTTAGCATTT GGACTCGCTT GATTGGCAAC GTGTAATCCT	720
20	TTATTAACCT TATTTAGGAA GTCGCTTAAA TCTGATACAT TGAGTTCACC TAATAAAAAAT	780
	ACCATTGAAG CCATATTAGA TAATAGCCCT GTATAAATAT CTTTATTAAG TTCAACTGCA	840
	AATTTATTTA TGATGACTTG ACGTCCTCGA ATTGCACCAT TTAAAGCATC TAATAGTTTT	900
25	GCATCATCTA ATGTTTTAAT AAGCTTGATT GCTTTTAATA TACTATCTTT ATTCTGCTGCA	960
	ATTGCTCTG TAACTTCATT TAACTTTCT AACTTAATTT GTTCTTCTGA TTTTCTAAG	1020
	CGTCTAATTT TAGAAGATAT TCTCTCAGCC ATTATTTATC CACCTGATTT CCCGGGAAAA	1080
30	CATAATCTGA ACGTTCCCAT TTTTCTGTA CTTGAACACT GTACTGCGGT TGACGTTTTT	1140
	TATTGACACG GAAATTATTA GGGTTCAACG GTGACTTACC ACGTTTCGTA ATTACCTCCA	1200
35	AACGACAGCT AGTACGTTTA TAAGATGGTG TATCCGTGTA TTGATCAACA TCACTaTTAG	1260
	TTAATAAGTT AATTGCACCT AGATCTCCAT TTTCCATCGC aTCaTTATTT AATGGAATAT	1320
	AGA <del>TT</del> CTTT ACCTTTAACA CGATCTGTCA CGTGAACCTG TAATACCGCT TCTCCTGtYt	1380
40	CAGAAATCAG CTTAACTTCT GCACCTTCAT GAATGCCTCT ATCTTCAGCA AGCTCTGGAG	1440
	AAATTTCAAC AAATGCACGT GGCACCTTGT ATTTAATCAT TGGTGTTTGA TAAGTCATAT	1500
	TACCTTCATG GAAGTGCTCT AACAAATCGAC CATTGTTTAC ATGAATATCA TAAATTTCAT	1560
45	CTTGCTTAAA GTAATTATCA AATGATAATG GGAATAATTT TGCTTTACCA TTATCAAAAT	1620
	TGAATCCTTC TAAGTATAGA ATAGGCTCAT CAGTACCATC AGGTTGTACT GGCCATTGTA	1680
	AACTATTGAA TCCTTCTAAA CGATCATAAC TTACCCAGC ATATAGAGGT GTTAAGCGTG	1740
50	CTACTTCATC CATAATTTCA CTAGGATGCT TGTAATTCCA ATCAAATCCT AATCTATTAG	1800
	CAATTGCTTG GAAAATTTTC CAGTCAGGTT TTkaATCACC AAGAGGTTCT AATGCTTGGT	1860
55		

EP 0 786 519 A2

	TTGCTGGCAA TACAACATCT GCGTATGTTG CTGTGAATGT TAAAAATTCA TCTTGGA	1980
	CCATGAAATC TAATTTTTCA AACGCAGCTT GTACAAAATT AATATTTGAA TCCACAATAC	2040
5	CCGTATCTTC ACCATATAAG TACAATGAGT GTACTTCTCC GTCATGTATA CCTTCTACCA	2100
	TTTCATGATT ATCTTTACCA GCTTTTGGAT TCAATTTAAC GCCATATTCT TTTTCAAATT	2160
10	TAGCGCGAAT ATCATCCGCT TCAATACTTT GATAACCAGT AATCTTATCA GGCATACTTC	2220
	CCATATCACT ACATCCTTGA ACATTATTAT GTCCACGTAA TGGATACGCA CCAGTACCAG	2280
	GACGACGATA ATTACCTGTT ACTAATAATA AGTTTGAAAT CGCTGTACTT GAGTCACTAC	2340
15	CAATGTCTTG TTGTGTAATA CCCATTGCCC AACAAATTAC AACAGATTCA GCTTTAGCAC	2400
	ATTCTTCAGC AAATTTAATC AATTCTGATT CAGGAATACC TGTGCTTCT TCAGCAAAAG	2460
	CCATTGTAAA TGTTTCTAAT GATTTGTAAT ATTCATCAAA ATCATCTACC CACTCATCAA	2520
20	TAAATGCTTT ATCGTGTAATA TCATGATCAA TAATATACTT AGTCACTGCA CTTAACCACG	2580
	CTAAATCCGT ACCTGGTTTA GGTGATAAAA AACGATCCGC ACGTTCTGCC ATTTTCATGTT	2640
	TTCTAATATC AAATACATGT ATTTTTTGAC CAAATAATTT TTGTGCACGT TTCATGCGTG	2700
25	ATGCGATAAC TGGATGAGCT TCGGCTGTAT TAGTACCTAT CAATACAGAC ATTGCCGCTT	2760
	TTTCTAAATC TTCAATACTA CCTGAGTCAC CGCCGTGTCC AACC GTTCTA AATAAGCCTT	2820
30	TTGTTGCAGG TGCTTGGCAA TATCTTGAAC AGTTATCAAC GTTATTTGTG CCAATAACTT	2880
	GTCTTGCTAA TTTTTCGATT AAATACGATT CTTCAATCGT CGCTTTAGAA GAAGAAATGA	2940
	ATGATAGTGC ATCTGGGCCA TGCTTTTCTT TAATAGCTGT AAAATTATCT GCAATGACGT	3000
35	TTAAAGCTTC ATCCCATCT ACTTCATGGA ACTCACCATT TTTCTTACT AGTGGTTTAG	3060
	TTAATCGTTG ATCTGAATTA ATATGTCCCC ATGAAAACCT ACCTTTAACA CAAGTCGCAA	3120
	TTTTATTGTC TGGAGAATCA TGTGATGGTT GTACTTTTAA AATTTCTCTA TCTTTAGTCC	3180
40	AAACTTCAAA TGAACAACCC ACACCACAAT AAGTACACAC TGTTTTAGTT TTCTTAATAC	3240
	GCTCTTTACG CATTTCTGCT TCTGAATCTG AGATTGCAAA TAGTGGACCA TAACCAGGTT	3300
	CTGCTTTTTT AGTTAAATCA ATCATTGCTG CTAATGAACC AGGTTCCGTA TCAGTCATAT	3360
45	AACCCGCATT ACCTTCCATA TTCACCTCCA TCATGGCATT ACATGGACAT ACCGTCGCAC	3420
	ATTGACCACA AGATACACAT GAAGACTCAT TAATCGGTAC ATCATTATCC CAAATAACAC	3480
	GTGGATGTTT ACGATCCCAA TCAATTCTAA TAGTTTCATT CACTTCGATA TCTTGACATG	3540
50	CTTCTACACA ACGCCACAT AAGATACATT GATTTGGATC ATAACGATAA AATGGGCCGT	3600
	AATCTTTTTT GTATGGCTTC TCTTTATATT CATACGTTTG ATGCTGAAGC CCCCATGCAT	3660
55		

	TATGCTTTTC TAAAATTCTGA TCAAGCGCTT CTTTTTGAGC ATCTTTCACA TCATTGTTCA	3780
	CAGTATTTAC AGTCATTGGA CGATCAATCA CCGTACTACA TGAACGTTCA ATTTTACCGT	3840
5	CAATCTCAAC AGTACATGTA TCACATGTTT GAATTGGTCC CATCGACTCG TTATAACAAA	3900
	TTGAAGGTAC AAAAGTATCT TGTGATTTAA TAAATTCAAG TAAATTCGTA CCTGGTTCTA	3960
	CAAGATAATC TTTTCCATCA AGTGTAACCA CCAAATGTTT TTGCATATTA CTCACCCCGT	4020
10	CTATATATAT TTTCCGTAAA TGACTTTTAA TAAATTGCTC ATATCCACCT AAAATAACGA	4080
	TGCCCCACAC ATCTTTCAGA TAGAATTAAT TTAATTGTAT TACTTTATGT ACTAGTTGTT	4140
15	AAGTAAATTT TTGTATTTTG CCTTTTACA ATCATTTTTA TTTGAAATAT TTTGCGCGAA	4200
	ATTAAATCAT CTTTTTGTTC AATTGAAAAT AATTATCATT ATTAGTTTTC CAATTATCTG	4260
	TTTCACGCTT TTTGCCATAT CTTTCACAAC CTTATTAATG ACAATATTTA ATAATCACCT	4320
20	CACCTAAAAA TCGTTATACT ATTTATAAAT ACCCTTTTTC TGAAAATTAA TAACCCAAGT	4380
	TTGATAAATA TCTACTATCA TTTAGAAGGT AATATTTATC TTTAAATTAA ATTTGTAATG	4440
	GATTAATTTA TAAAAATCAA ATCAGGCATT AAATAAAATA GCCCATAAAT ACAAAGTGTT	4500
25	ATCACCTTCT ATTTACGGGC TATTAGTTCT ATTCGTTATT CTATTTACAG ATCATTCTAT	4560
	CTAATTAATT TGTGTACAAT TTTGATAACT TATTTTCCCT TAGTTTACTA CTCTAGATTA	4620
	TCTTTTAATA ACTTAGTACT TTCAGCTTTT GACTGCTCAC TAGGAATGAA GTAGTACAAT	4680
30	CCGTCACTTT GAATGCCGCC TTGACCACTC AATTGATGTT TATTAATCGT GTCATTAGCA	4740
	TCTTTATAAT TGCTTCTAAT CGTATTCAAA TCACCTAATG TTAAATCTGT TTTAACATTA	4800
35	TTTTGAATTT CATTCAATAG ACTATTAAAA TGTGTAATCG ATGATGGGCT TGCAATCTTA	4860
	TTGGCCATCG CTTCAAGCAC AATTGCTGA CGTTGTTGTC GACCAAAGTC ACCACCAGCA	4920
	CCTTCTTCTT TACGACTTCT AATAAACTTC AATGCTTGAT CACCATTAC ATGTGTCTGC	4980
40	TGTCCTTTTG TAAACGAAC ACCATCAACA GTGAATGTAT CATTACTTAC TACATCAACA	5040
	CCGCCGATGC TATCTATCAT ATTATGCAAA CCATCCATAT CGATTGTCGC ATAATGATCA	5100
	ATTGGCACAT TCATTAATTT TTCAAGTGAT TTAACAGCCA TATTTGGTCC ACCATATGCA	5160
45	TAGGCATGTG CAATTTTTC AGTAGTACCA CGGCCAACAA TTTCCGCTCT TGTATCACGC	5220
	GGTATACTTA CTATTTTCACT TTTCTTCGTT TTAGGGTTGA TAGATAAAAT CATAATACTA	5280
	TCCTACGCT CTCCGCCACC CTTTTTCTTA CGATCAGCAT CTGAATCGAC ACCAAATAAA	5340
50	GCGATTGTGA ATGGATCACC ATCGTTTAAA CTCACTTTTT TATCTCTTAA TTCTGAATGA	5400
	TTGCGATCTA ACGGATTGTG TATCTTATTA CCAGTAATAA AAATTTTAGC AGCTACATAC	5460
55		

## EP 0 786 519 A2

	GGTAGGCTCA TTTTACTTTT AGACGAACGT TTCAATCCCA CCACTCCTTT ACTATTCCTT	5580
	ACATACTTTG TCTGTTTTCT CTATTTATTA TATAGTAAAA TAATTTTTTT ACTATACTTC	5640
5	TGTAGACGTA TAACTATTTT TTATCATT TTATCTCTAG AGAATATCTA TCTGTATTTT	5700
	TGATAACCAC CATTTCATT TAAAATTTTA AGTACCGTTT CATGACATGC TTTATTACTT	5760
	ATAATAAAAG GTGCACCCTT TAAATGATCA ATTGCCTTAC CATCTAAAGT CGTCATTTTT	5820
10	AGATTCAATA GTTCTGCAAA TAAAACTGT GCAGCAATGT CCCAAGGTTT AGGATTTGTA	5880
	TTAATATGTG CCCCAAATTG ACCTTTTGCC ACTCGCATAG AATCTAATCC GCAAGCACCA	5940
	ACTAAACGAT AACTAAATGA GCGTCAAAT AAATCTTGCA CCGTATCTAG ATTCATCACT	6000
15	TGTGCATTAA ACGATATAAT AGCGTCTTCC AATTTTAACG ATGGTGGTTC TTCCATCTTA	6060
	ATTCCATTAC AAAAAGCACC TTCTCCTCGT ATTGCTTTAT AAAGCTTTT ATGCGGATAA	6120
20	TCATATACGT ACGATAACAT TGGTTTACCT TCATAAAAAT ACGCCAATAT AATACAATAA	6180
	TCTTCTTGCT GTTTTACTAA ATTGGCAGTT CCATCAATGG GATCCATAAT CCATAAATGA	6240
	TTAATTTTCA TCGTAATCAT TTCATTACTT TTTTCTTCCG CTAATAGTTG GTGTTCCGGA	6300
25	AAATGTGTTG CTAAAAATTG TTGGAATTGT TGTGAATCT GTTTATCTAC ATTTGTAACT	6360
	AAATCAAATC GATGACGCTT AGTTTCTGTA GTCATTTCCA TAATTAATTG CGGAATAACA	6420
	TTGTCTATTT GTTTCAACCA CGAACATATT AACTTATCTA TTTGCTGTAA TGTMTTATCT	6480
30	GTCATTTTCG CCACCACTTC TCATATCATT ATCATTTTAT TATTACCCTA TATTAAAAGA	6540
	ATCAACAATA CAACTGAAGA CTTCTTCATT TTATGCATAA AAAAATCGGC TAGTCACGTG	6600
	CTAGCCGACA AATAGAAAGG AAAGTAAGTA ATAAATATTG AAGATGTTGT GATGTAACCT	6660
35	GAACGATTAA AAGCTATCTG TTATATAGCT CTACCCCTTT GTTTAATCGC TCCCCCTGTT	6720
	ACAAGTAATA TCATAGCACA ATCTTTTTTA AAATGTAAGC GTTTTCCACA AAATTTTTAC	6780
40	GATTTTTTTA AAAAGATATT GAAAATGTCC TCATTGTCAC TCTTATGTTA TACTTTGTGT	6840
	AATATATCAT CTTTATAGGAG GTGGCTGTCA TGAATAAAGC TGAAAGGCAA AATTTAATAA	6900
	TTACTGCAAT TCAACAAAAT AAAAAATGA CCGCTTTAGA ATTAGCTAAA TATTGCAACG	6960
45	TATCCAAACG CACAATTTTA AGAGATATTG ATGATTTAGA AAATCAAGGT GTTAAAATTT	7020
	ATGCGCATTG TGGGAAAAAT GGTGTTACC AAATACAACA AGCACAATCT AAAATTGCAT	7080
	TAAACTTATC TGAAACACAA TTATCAGCCT TATTTTTAGT GCTTAATGAA AGTCAGTCGT	7140
50	ACTCGACATT ACCATATAAA AGCGAAATCA ACGCAATTAT AAAACAATGT TTAAGTCTTC	7200
	CACAAACACG CTTAAGAAAA TTGCTTAAAC GCATGGACTT TTATATTAAA TTTGATGACA	7260
55		

	ATGTGATGTT AGTAGATCAT AGGGTTGATG ATAATATTAA AGCTGAAAAC GTTATATTTA	7380
	TTGGCCTTTT GTGTAAACAT GGACATTGGC ATGCAGTCAT TTATGACATT GCTCAAGACA	7440
5	AAACTGCCGA ACTCGAAATT GAAAATATTA TAGATATTTT GTATTCATTG GGTAAGACGA	7500
	TTCAAACCAG AGACATATCC ATTGATAACT ATCATCAATT TTTAAACCCC ATCGATTCCCT	7560
	AAAAACAGC AGTAAGATGA TTTTCAATTA GAAAATATCT TGCTGCTGTT CTCTATTTAT	7620
10	ACAATACTTC GTATTGAATG GnTTCGCTTT CCTAGGGTGC CGTCTCAGCC TTGGTCTTCG	7680
	ACTGGCACTG CTCCCTCAGG AGTCTCGCCA TTAATACTAC GTATTAACAT GTAATTTTAC	7740
15	TTTGAAATAC TTAaaaaaAT AAAACACTTT GCCCACTTA CACTACCAAT AGAAACTGCT	7800
	GTTAGAATTC CTCAAATGA TATTTGCGGA TATGTTAATG AAATTGTTAA AAAGATAGCT	7860
	GATAGCGAAT TCGATGAATT CAGACATCAT CGTGGCGCAA CATCCTATCA TCTAAAAATG	7920
20	ATGTTAAAAA TCACCTCATA TTCATATACT CAATCTGAAT TTTCTGGCCG TAGAATAGAA	7980
	AAATTACTTC ATAACAGTAT TCGAATGATG TGGTTAGCTC AAGATCAAAC ACCTTCTTAT	8040
	AAACTATTA ATCTTTTTAG AGTGAATCCT AATACTGATG CGCTAATTGA ATCTTTATTT	8100
25	ATTCAGTTTC ATAATAAAAT GCATATCAAA AAAGCTGATT TCTATCAAAT AATTAATAGA	8160
	AATCAGCTTT TTTCaTTGCC TAAAACTTA ATGTCCCGAC CTCTTTATCT ACGCATAAAT	8220
	ACTTATTACT GATATAACGA AAGAAACAAA ATTATTTGCT ATATGTAATG CAATTGTTGA	8280
30	ACCTAGGTTT CTTCAGATT TTAAATAAGT GAAAACTAAT ATGATGGATA GTATGAGATA	8340
	TGGACCAAAC TCAAACGGCG ACTTTGCATC AGTCACATGA ATAAATGCAA ATAAGAACAC	8400
	CGAAACAATA CTCATAGCTA TAAAATTAAA CTTCTTACCT AATTCTCAA TTAaaATATG	8460
35	TCTAAATACG ATTTCTTCAA CTATTGGACC TACAATCACA ATTAATAAGA ATGCTACAGG	8520
	TAAaAATGCA GGCACCTCAA ACATTTTATT TAGCTCAAGT TCATTGGCTG TTtCACTATA	8580
40	TTGCAATGT TTAGGTAGAA ACTGTGTCAT ATATTCAATG GTATAAATTA AGATGAGAGC	8640
	AATAATATAC GTTATTGACA ATCTAAGCCA ATATTTTTTG ATATACGCAA AACCAGCTCG	8700
	AAGCCTTGAT GGCACTACTT TTAAATGAAA TAAATAAAAT GCGCCAATCC CAATCGTATA	8760
45	TGCTAAAGCT TGTGTGATAG TCGCTACAAA TATCAGATTA CTATCGATTT CATAATAACC	8820
	AAACAAAATT GGTCTATGT AAGCTGCAAT TGTGAGTGCA TAAATATAA CACCTATAAT	8880
	TGGAATTATA AGCAAATCTC TCCATGCTAT ATCTTTAAAC GTGTATTTCT TTTTTTCATT	8940
50	TTCCaCTGTT ATATCCtTTC CTGTTTAATA ATTGATTTTT GGAGGTACTT CTACATGATA	9000
	AACGAAACTA AGTATATGAG ACAACAAATT ACTAATTTGA TTCAAATCAT TGATACGATT	9060

55



ATAGTTACTA ATGAATTGAA TAAGTTCAAA GGCTTTGAAA CATCATATAT AATAAACGAA 9180  
 AATCAAGTTT CCTATTATGA AATTATAACA CTACTIONAATA AACGTCCCCT CgACAAGTCG 9240  
 5 ACTATGGTAA CAAAATTCAA TATCTTAATT TTTATCATAC AGAACTATCT AACGCATTAT 9300  
 TTGCAATTAA ATTTGCCCAT TAACCTATTT TTCATAAAAT GTCATTTAAA CAAGTTATTT 9360  
 ATTAAAATTC ACTTTATTAC ATAAATTATA CAATTArAAA GTTTCCTCAA ATTGTAAAGA 9420  
 10 TGCATTAATC GAGTTATAAT CATAATGATT AAGATGGT 9458

## (2) INFORMATION FOR SEQ ID NO: 115:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 910 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

AnGCGTATCA TGTCACGCAT TTTAACTACT TCTTTACCAC AAGATTATAC AGTCACATTA 60  
 25 GTTGATCGTA TGCCATTTC A TGGATTGAAA CCAGAATTTT ATGCTTTAGC TGCGGGCAGG 120  
 AAATCAGATA AAGATGTTTG TATGAAATTC CCTAATCATC CACAAGTGAA TACAGTTTAT 180  
 GGTGAAATTA ACGACATAGA TTTAGATGCT CAAATTGTCT CAGTCGGTAA TTCTAAAT 240  
 30 GATTATGATG AGCTAATCAT TGGTTTAGGA TGTGAAGATA AATATCATAA CGTTCCAGGA 300  
 GCCGAAGAAT ATACACATAG TATTCAAACA CTCTCAAAGG CTCGGGATAC TTTCCATAGT 360  
 35 ATTAGTGAAC TACCAGAAGG TGCTAAAGTC GGTATCGTTG GTGCTGGATT AAGCGGCATA 420  
 GAACTTGCCA GCGAATTAAG AGAAAGTAGA TCAGACTTGG AAATATATCT TTATGACCGT 480  
 GGGC<sup>~</sup>CGCGAA TTTTAAGAAA TTTTCCAGAA AAATTAAGTA AGTATGTTGC GAAATGGTTC 540  
 40 GCCAAAAATA ATGTTACCGT TGTTCCAAAT TCAAATATTA ATAAAGTTGA ACCTGGTAAA 600  
 ATATATAACT GTGATGAACC TAAAGATATT GATTTAGTTG TATGGACAGC AGGAATTCAA 660  
 CCTGTTGAAG TTGTTGTA CTTGCCGATT GATATAAATA GTAATGGACG CGTGATAGTT 720  
 45 AACCAGTATC ATCAAGTACC AACATATCGT AACGTCTATG TAGTTGGTGA TTGTGCTGAT 780  
 TTACCACATG CGCCAAGTGC TCAGTTAGCC GAAGTTCAAG GTGATCAAAT TGCCGATGTG 840  
 50 CTTAAAAAGC AATGGCTAAA TGAACCATTA CCTGACAAAA TGCCGGAAGT AAAGGTACAA 900  
 GGTATCGTTG 910

## (2) INFORMATION FOR SEQ ID NO: 116:

55

(A) LENGTH: 10182 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

10	TTTTTGATTC AAAGTGGTGA TTTAACAAGC ATTTTAAATA GCAATGATTT GAAAGTCACA	60
	CATGATCCTA CCACTGATTA TTATAATTTA TCTGGTAAGT TGTCGAACGA TAATCCAAAC	120
	GTAAACAAT TAAAACGTAG ATATAATATT CCTAAAAACG CATCAACAAA GGTGGAATTA	180
15	AAGGGAATGA GTGATTTAAA AGGCAATAAT CATCAAGATC AGAAACTTTA TTTTATTTTT	240
	TCAAGTCCTG GAAAAGACCA AATCATTTAT AAAGAAAGCC TTAATTATAA TAAAATAAGT	300
	GAACATTAAT ACTTATGCTG TAATTATAGA AACATCCAAA TCATCTATTA NAATCCTATA	360
20	TTATAAAAnC ACCTCACATA ACTCGTTCAA CTGTACCAA CCACATTACA TTAGATTTTA	420
	GGCTAACTAT TGTGATGTAC ATCAAAAACG AATTTGTGAG GCGTTGTATA TTTTACAAAG	480
25	GTGACTAGCG TTTCGTATAG CATTTCCAAC ATTACTACAC TCAAGCGTCA CGCTAAAGTT	540
	CGAAATCGAA TCCTTTCATT CAACAAAAGC TCATATCCAC TACAACTTC ATATCAAGCG	600
	TATAAACTAT CTTGTGATAC TATCTCGATC ATATCTATAG TATGCATTTG TGTTCCGTTT	660
30	CACTGAAGTA TATGTATCAT CAGTTAAGTA TAAACCGTCA TCCTTCAATG TTAATTGATA	720
	AGCATATTTT CGTGCTAACC AGGCAATATC TATATAATTT TCTCCTGCGT TTTTATAACT	780
	TCTTAAATCT TCAATATGTG CACTAACTTC AGGGaAAATG ATTCTAACAA CACTTTCATC	840
35	AACCCAATAT TTGTCATGCA TCCATCGCAC TTGATCTGCC AATAAAGGTA ACTGCACATC	900
	ATTGAAATAT AGACGAAAGC CGTCACTATC ATACATTTGC CGATATGGTA ATGGCTGTTT	960
	TCTAACTACT AACACCTCGC CACCCATTAC GGTGCCTTCT CTAGTATCAT CACTTCCACC	1020
40	CGAAGCTTCA TACGTTGTTG GGTCAACCTG TAGTCCATGT ACATCTCCAA TATAAGCATC	1080
	TGGTTTATGT TCCATTGCAT GTCCATGTGC AATCAATGCT AATATTGTAG ATTGTGAAAA	1140
45	TTGAGGCTCC CATTCAATGC GATTAGGATG GCTACTATAA ATTCTAGGTT CATCTATAGC	1200
	CTGCTGAATA TCCATGCCAA AACTAATAC ATTGATTAAT GTTTGCGCAA CACTAGCAAT	1260
	GATACTTATG GCACCAGGTG CACCTACTGT TAATATTGGC TTCCCGTGAT ACATCACAAT	1320
50	CGTTGGAGCC ATGTTACTTA GTGGTCGTTT ATATGGTGCA ATTTTCGTTAA TACCACCATC	1380
	TACTACATCA AAGCCATCCA TTGTCGTATT CAATAACACA CCGTAGCCTG GAATCGTGAT	1440
	ACCTGAACCA TAAATCATAC CAATTGATGT CGTAAATGAA GCAATATTAC CTTCTTTATC	1500

55

EP 0 786 519 A2

	ATCAGACACA ACACCATGCT CTATATCAAT ATTTGCTTTA TTGCTATCAA TGAGCGTACT	1620
	GCGTGCTTTT AAATAATCAT CATCAATTAA TGACTGTACA GGCACCTCAT GAAAATTATC	1680
5	ATCCGCCAAG TATTGCGCAC GATCACTATA TGCTAAATGC ATCGCTTGTA TCAAATGATG	1740
	CAAGTAATCA ACAGATCTTG GACCCATAGA TGGTAAATCG ACATGTTCTA ATAACTTCAA	1800
10	TATTTGAATT ACCGTGATAC CGCCAGAACT AGATGGTCCC ATTGaaATAA TGTCATAGTC	1860
	TTTAAATGTT GCACTGATTG GCGCTTTAAT CTGAATGTCA TATTTGGCTA GATCCTCTAA	1920
	AGTGATTGTC CCACCACATG CTTTGACAAC ATTGACTAAT TGTTCGCAA TGTCACCTTT	1980
15	ATAAAATGCA TTAAACCTTT GTTCTCTTAA TATTTGAAAT GTCTTACCTA ATTCGGGTTG	2040
	TACAATCCAA TCACCTTCAC GCCAATATTG ATTTTCATGC GTAAATACTT GTGCCGTTTC	2100
	ATGATACTTT GTCAATCGTG CGTGTGTCTG GCGCGAATAT TTTTCAGTAG CCCAATTGGC	2160
20	TGCATGACCT TCAATGGCTA GTTCAATTGC AGGATTAATT AAATCTTCCA ATGACAATTT	2220
	AGCATAACGC TTGTGAATAT AATCAAACAG CTTTGGAATT GCTGGCACAG CGACAGTTTT	2280
	ACCATGTGTA GTCATATCAA AAAATGATTT ATATTCGCCT GAATCATCTA GATAAAATG	2340
25	TTTGTCTACA TGTTCAGGTG CTGTCTCAG TGATCAAAC GCAGTTATAC TGCCAGTACT	2400
	TTGCTCATAA TATAGCAAAT ACCCGCCACC ACCAATACCT GATGCAAATG GTTCTACCAC	2460
30	ATTCAATGCC AGTTGAATTG CAATCACTGC ATCCATGGCG TTGCCACCTT GATCTAATAC	2520
	ATCCTTACCA ATTTTAGCCG CAAGAGGATG TGATACGGAA ATTAACCTT CTTTAGATGT	2580
	TTTTGTCTGT TTGTCAATTA AGTTAATGAC CATACTATAT CCTCCTACTT TCTGTAAAT	2640
35	ATTTAAAACA TTATTGATTA ATGGCTTTTT CTACTTTTTT TAAATCTTGA CGTTGCTCGT	2700
	TACCAATATC GACAAGTGGT GTAATCGGTG ATGCAATTTT AAATTTATCG CCACGATAAA	2760
	ACTTAATAAA TTGATCCTGA TCTATCGCAT TAACTACTGC TTGTCTCAAG TTTGGATGCG	2820
40	TCTTAAATAT ACCTTTTTTA ATATTTAGCA TTAAAAAGAC TGACTTGCGT CCATTTTTGC	2880
	GAATAATGCT TAAATTTTTA TCCGACTTAA TTAAATCAAA ATGTTTTTGA TTCACATCTG	2940
	CCAACATATC AATTGAATGA TTTCTAAGTT CTGACAATGC ATTATTCGGG TCACCATTAA	3000
45	ACTTCAATGT AATATTTTTA ATTTTAGCTG GTCCATAACT ACCTTTTTCT GTTTCGTTGA	3060
	ATCCTGGATT ACGTTGAAAC GTTGCTTGAT ATGCATTTTT CTGTGTCATA ATGTATGCGC	3120
50	CACTTGATA CAGCGCATTT TTCCCATCTG AATTGTCAGG AATTGTACTG CTATCCCCAT	3180
	ATCCTTTTGG ATATTCTTGA TTTACTTGAT TAACAAATTT TTTAGATAAA ATGCCTGCCG	3240
	AAGAGTGTGT TAAGTAATTT ACCTCTCGAG GCATCGATTG ATCTGTCGTA ATTTTAACAA	3300
55		

	TATAAGCTTT	AATCAACTTA	TCATAGATTG	ATTTATCGTC	CTTGTCTTTC	TCTTTACGCA	3420
	ACTGATCGAT	GTCCTCATCT	TTTAATATCT	TGATGTCATT	TATATGTTTG	TGCATATTGT	3480
5	AAGTATTATT	GTTAGGCACA	GACTTTTTTAT	CACGTGCTCT	ATCTAAAGAA	AACTTAACAT	3540
	CTTCAGCCGA	TACACGCTCT	CCAGTATTAC	GTGCTTGTC	ATTGACCACT	TTCGCAAAAT	3600
	AATCATCATC	TCTTAACAAG	AAATAAAATG	CTTTATTGTC	CTTATTACCA	GCATAATCAT	3660
10	GACTTAACGA	ACCTTTCGTT	GTTAAATGAT	CATTTTCATC	TAATAATAAT	AACCTTGTGT	3720
	ACATATTTCAT	ATTAATTGAA	TATACTGACG	GCGCAATTGA	ACGTATTGGA	TCCAATGTAG	3780
15	GAATTTACC	ATCTTGTTGT	GTCAATCACA	GTGGCCGCGT	ATCTCGTTCT	CTACTATTGT	3840
	TGTAATCAAA	TTGTTGCCAT	ATTAATGCAC	GTGAATTTGG	CAATCCAACA	CTATTTTAT	3900
	CTAACACTTT	ATTGTCATAT	ACTAAATTCT	TTTTTGATCC	ATATAAAGGC	GCCATATACC	3960
20	CTTTATCAAA	TACAACTTCA	TCTTCAATTT	GCTTATATGT	TTGTTTAAAC	TCTGCTTCAT	4020
	TTTGAGTAGA	AGCTTTATTT	AACAACGGT	CTACATGTTT	ATCTTTCAAT	AACTATTTG	4080
	ATCCTGTAGA	ACTAAATAAT	GCCGTCATAG	CATAGTTCGG	GTCACCAAAC	ACTGTCATCC	4140
25	AGTCATCAAT	TTGGATATCA	TAATTGCCGG	CTTGACGTG	TGTACGATAG	CTACCATAAT	4200
	CTGGTTGGAT	ATTCATCTTC	ACGTTAAATC	CTGCATTTTC	CAATTGATCT	TTAACGATAT	4260
	TCATATCATT	TTCATAACTT	GCTTGTCTTA	GGAAATGTAT	TGTTGGTCGC	TCGCCTTTCA	4320
30	CTTCAACTTT	CGATGACTTT	TGAGCCACTT	CTGATTTTCG	AGGGACACCA	CAACCACTTA	4380
	ATACCAACGC	TAAAACTATA	ATTGCGATAC	TAATGATTTT	CTTCACATCT	ATCCCTACCT	4440
35	TTTAAATGAA	TTCTTGGATC	TAGTGCATCA	CGCACTGCAT	CACCTATAAA	ATTAAATGCT	4500
	AAAACGACGA	ACATAATACA	AACACCAGGT	ACAATAGCTA	AATTACTGTG	CGTTTCCAAG	4560
	TAGTTACTAC	CGGTACGTAA	AATGTTGCCC	CATTCAGCTA	CATCAGGTGC	AACACCAAGT	4620
40	CCTAGGAAAC	TTAAACTACT	TGTTGTTAAT	ACAACCACAC	CTATATTTAA	TGAAAAACGT	4680
	ACAATCATAG	GCGCAATCGC	ATTCCGTAAA	ATATAACGCC	ATATGATATT	CCAAGTGTTT	4740
	TCACCAGTGA	TACGTGCTGC	ATCTACATAT	TCCATGCGTT	TAATTTCTAA	AACACTGGCA	4800
45	CGCATTGTCC	GTGCAAATGA	TGGTATATTA	CCGATACTTA	AAGCAATAAT	TAAATTTGGA	4860
	ATACTTGCTC	CAATGATGC	AATAATTGCC	ACCGCTAACA	ATAATGATGG	AATTGCAAAC	4920
	ACTACATCTA	AAATTCGCAT	TATTAAATTA	TCAATATGAT	TAAAATAACC	TGCGATAGTG	4980
50	CCTAGTAACA	CACCAAAAAT	AAGTGAATA	ACTACTGAAA	TAATTGAAAT	TGAAAATGTC	5040
	AGCTTCGTTT	CTACAACTAC	GCGTGTAAT	AAGTCTCTAC	CGAAATCATC	AGTACCAAAC	5100
55							

	GTATCAAATG TAAATTGTGA CACAATTGAT AATGTCAGCA TGTAGACTAA AATAAGTAAC	5220
	CCGATAATCG CAATACGATG TCTAGTAGTT TTTCGTATAA ACGATTCCCA CCCGTTATAA	5280
5	CTATGTATTT GCGATGTACG TTGGTAAAGT CTAATACTTA CAAACATTAA TAATGTAAAT	5340
	ACGTTGCCTG TTAATGTCAT CAACAATAAC AACACTTCGA CGATACGTCG CCATAGGTCA	5400
	TGATGCTTCC ATGTTTGTTC CGTTGTAAAA ATAATAATTA AAATGATGGT TAAAACGATT	5460
10	AGCAATGTTT CAGCAATATA GAACGTATCG GCCACATAAC CTTTAAAAAG ATTTAATGCA	5520
	CTCGTTAATA TAACTAAAAT ATAAGTTGCT ATGGCGTAAC TTGCGAATAA TTTTAAGGAA	5580
15	GCTATCTTTG AATTAAGTTG TGCCATATGC CTCACTTCCT TTCGTTGATT TCACTACGTA	5640
	ATTTTGGATC GATTAAAGCA TAAATATAT CAATAATTAA GTTTGCTAAA GATATTACAA	5700
	TTGATATATA TACGACCCCA CCCATGACTG CTGGAATATC AGGTATTAGT TGTTTTTGGA	5760
20	CGATATAACG CCCGATACCA TTAATGTAA ATACTTGTTT CGTCACTGCT GAACCGCCTA	5820
	GTAACCTGTC CACTAGAAGA CCAACTAACG TTACAATTGG AATAATGGCA TTTTCAAAA	5880
	TATGTTTAAT AACAACTTGT GTCGTCGATA ATCCTTTTGC ATAAGCAGTT AAAACATAAT	5940
25	CGctGCGCAT TACTTCAAGT ACAGAAGACC TTGTCATACG CGTGATAGAA GCAGCAATAC	6000
	TTGTTCCAAT GACAAGTACA GGTAATCA ACGATATTGG ATGTTCTGGC ATATAAGATG	6060
	GTGGCAAAAT ATCCAATTC AATGAGAACG CTAAATGAA TAATAGCCCT TGCCAGAAAC	6120
30	TTGGAATAGA TAAACCAATT AATGCAATTA TCATTAACGT GATATCAAGC CAACTATTTT	6180
	GCTTCATCGC ACTGATAATA CCAATTGGTA TTGCAATAAT TAATGCCACC ATTAGCGCTA	6240
35	ATACTGCGAC AATTATTGTA ATTGGAATTC TTTCGCCAAC TGCTTTAGTC ACAACCTCAT	6300
	TCCCTTTGTA AGTCGTACCT AAGTCAAAGG TAAAAACACC CTTGATGGTA TCCCACAATT	6360
	GAATBAAATA AGGTTCGTTA AGATGATGTA ATACATTGAA TTGATGTATC TGTGCCTTTG	6420
40	TTGCATTTTG TCCCAGTATG CTATAAGCCG CATCAAGCGG TGAAAAATAC AGAATGGTAA	6480
	ACACACTGAC AATAACACCA ATGATGACAA TCACAGCCAT GACAATTCGT TCAAAAATAT	6540
	ATCTAACTAA TGGCTGTAAA TAAAAAGTCA ATAAGATGAA CATCGGCAAG GCCAATATCA	6600
45	CTTTGATCAT GATGAACCTA TGAAATAATA CATTTTCAAA GTATGTTGAA AAATGTGCTT	6660
	GTTCAATATT CTTTGAACCT GTATTAGAAC TTTGTGCCTT GAATATTTTT AATGCTTCTT	6720
	TATGTATTTG TGTGGATGAC TTTTGCTGCG ATAAATATTT ATATTTTGA TGTAACGCCT	6780
50	GTTCAATTTT TGAAATTTCA GAATTATTAG CGTAAAAATT TTTCCTCTTA GCAGAAAAGA	6840
	AAAACCTTTAT CACTGCATAT AAAAATATTG GCAAGCTTAA TACCGATAAT ACAAACCTTG	6900
55		

	CTTGTA AAAAT AATCTTGAGT AGATTACTAT GATATACAAA AGTATAGAAT AAATTTACAC	7020
	ATTTGTGaAT AGGGAGGCAC AACATCATGT CAAATTTATT AGAAGTCAAC AGTCTGAATG	7080
5	TACAATTCAA TTATGATGAA ACTACAGTTC AAGCGGTAAA AAACGTCTCT TTCGAATTAC	7140
	GAAAAAACA TATCCTAGGT ATTGTTGGTG AATCAGGATC AGGAAAAAGT ATTACCGCTA	7200
	AATCTATTTT AGGGCTACTA CCAGATTATC CAGATCACAC ATTAACAGGA GAAATTATTT	7260
10	TTAATGGGCA ATCGTTAAAT AATTTATCAA CTTCAGCGTT ACAACAAATT CGAGGTAAGG	7320
	ATATTTCAAT GATTTTCAA GATCCACTCT CTTCGTTGAA TCCAAGATTA ACGATTGGCA	7380
15	AACAAATTAC AGAAGTAATA TTTCAACATA AACGTGTATC TAAATCTGAA GCAAAGTCGA	7440
	TGACAATAGA CATTTTAGAA AAAGTAGGTA TAAACATGC AACTCGACAA TTTGATGCTT	7500
	ATCCACATGA ACTTCTGGT GGTATGCGTC AACGTGTCAT GATAGCAATG GCATTGATTT	7560
20	TAAAGCCACA AATTTTAATC GCAGATGAaC CAACAACGGC ATTAGATGCC AGTACACAAA	7620
	ATCAATTACT GCAGTTAATG AAGTCCCTTT ATGAGTACAC AGAAACATCT ATTATTTTTA	7680
	TCACTCACGA TTTAGGCGCT GTGTATCAAT TTTGCGACGA TGTGATTGTA ATGAAAGATG	7740
25	GAAGTGTCTG TGAAAGTGGC ACGGTTGAAA GTATTTTTAA ATCGCCACAA CATACTATA	7800
	CAAAACGCTT AATAGATGCG ATTCCTGATA TTCATCAAAC GCGTCCGCCA AGACCGTTAA	7860
	ACAATGATAT TTTATTAAAA TTCGATCGCG TGAGyGgGAT TACACATCAC CGAGTGGCAG	7920
30	CCTATACCGA GCAGTTAATG ATATTAACCT GGCTATTAGA AAAGGCGAAA CATTAGGCAT	7980
	TGTCGGTGAA TCAGGGTCAG GGAAATCGAC ATTAGCTAAG ACGGTCGTCG GTCTAAAGGA	8040
35	AGTGTcAGAA GGCTTTATTT GGTATAACGA ATTACCATTA AGTTTATTTA AAGATGATGA	8100
	ATTGAAATCT TTACGACAAG AGATACAAAT GATTTTTCAA GATCCATTcG CATCTATTAA	8160
	TCCAAgATTT AAAGTCATTG ATGTGATTAA ACGACCACTA ATCATTcATG GGAAAGTCAA	8220
40	AGATaATGAT GACATTATTA AAAGTGTCTG ATCGTTGTTA GAAAAGGTTG GCCTAGATCA	8280
	AACTTTCTTA TATCGCTATC CACACGAATT ATCTGGTGGG CAACGTcAGC GTGTAAGTAT	8340
	CGCGAGAGCA CTTGCTGTTG AACCTAAAGT GATTGTTTGC GACGAGGCAG TGTCCGCTTT	8400
45	AGACGTTTCA ATTCAAAAAG ATATCATCGA GTTATTAAAA CAATTACAGT TAGACTTCGG	8460
	CATCATTAT TTATTCATCA CACATGACAT GGGTGTATC AATGAAATAT GTGATCGCGT	8520
	TGCAGTTATG AAAAATGGCG AAATCGTTGA ACTGAATAAC ACAGAAGATA TTATCAAACA	8580
50	TCCGCAGTCA GACTATGCAA AGCAACTTAT TTCAGAAGTA GCAGTTATTG CTAAATAAAA	8640
	GTCATGCGTT GTGCAACTTT ATCACTGTAT GGTCTGAAAT AAATTGCGCG ACTTCTGATG	8700
55		

TATCAAGTTT TAGGTGCTTT GCCATGATTT AAGAGTCACC CCCATACTTT GGGCATTTTA 8820  
 ACGCCAGAAT AAATCCCCCG CCACTATGTG AAGTGTGGGG GATTATTTAT ATTTTATTAG 8880  
 5 AATATTCAGA TTTTGTAGTG TGTCAACTTA GCTTAGTCAA TGTATATTTA ACGTCACTTA 8940  
 CTCTTTTCTT TTCATAATTA ACACATTCAA ATAAACTTTG ATCAAAAAAC ACAAGTTAA 9000  
 AAGTACCATC TTGTAATATG CTCTCATACA TTATCCCGTC ATATTAAAGG CTTCGAATAT 9060  
 10 AATCAGCTAA ATATTGAAAT GGCAAATAAT CTATTCCTTG TTCATCGCTT GGATTTGTGA 9120  
 TTCCTTTATG AATCTTTTTT AATGTTTGGT AATTTACAAA ATACTTTCTA AATCCATCAT 9180  
 15 CGCCAGCTTT GATTGCATTA CTAGTTAAAT TAGTTAAATT CGCAATTTTC AATTTCTCTT 9240  
 TTGTCACGTT TTTTGTAAAC TTAACCTTAC CTATATAAAT AATGTCATTA TGCTTAGGTT 9300  
 TAACTTCTTC TATACTGACC TGTCTTTTTG TACTAAGGTA TAATACGCTT ATCCATTTAG 9360  
 20 AATTCAATCT TCCTGCCGTT GCAAATCCCT TTGGTGGTGA CATTAGTTCA CTTTCTCTG 9420  
 TAATGAACTT AACTATTCTA GATCTATATA ATGGTTCAAA TCTTCTCTA AATCCTCAA 9480  
 TACTATAGTA ATTAGTAGTG ATATCGAGAA AGACGCTAA ATTCTCTAAA TTGATCATAT 9540  
 25 TTTTATGAAA TCTATTTTAA TACTTCAAGC TCTCAAAAA TCCATCCCAG TCATTATTTG 9600  
 CTACAATTAG ATTTTATTTT GTATATTTT TATCGTTTAT GATTTTAGCG CCTACTAAAT 9660  
 CTTCCAACAC TCGTCTATCT AAATTTTCAT CATCTTTAAA AAGTTCATTT AAAATACAAC 9720  
 30 TTATTTGAGC TTCCTCAACA TTAAATATAC TCCAGTCGTC TTTTAATGCT ATTTCAATCT 9780  
 TTTTACCTTC TTTTGGGCTA AAAGTATCTG GTAAATTTAT ACTAATATCA TATAATTCTA 9840  
 ATGCTGGTCT TAAATAATCT CTAATAAGTT CTAATTTATC TATGTCCTTA GTCGTATCAA 9900  
 ATATTTTAAC ACCAAGATGA TTGTTATCAA TATCACAATT GTCAAATTTG CTAATTATCA 9960  
 TTTGCAATGA TTTCTACGAT TTCAGTATTA TTAACACATT TTTCACATAT TTTCATTTTG 10020  
 40 AGACTCCAAG TATCTATTCA TAATTTCTAG GTGATGCATG ATAGATAACC TTTTAATTAA 10080  
 ACCTAATCCT GGATaCTTAT TATTTTCATT TAATCTTCA AATTGTCCCA AGCGCATAAG 10140  
 ATCTATTTTT AATATCTAAG TTTTGTGACC ATGTTACTAA TT 10182

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3491 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

	AACTCAGGCA ATTGAAACAG CATTAGGTGC TTCATTACAA CATGTCATTG TAGATTGAGA	60
	AAAAGATGGA CGCCAGGCTA TTCAATTTTT AAAAGAACGT AATTTAGGTC GTGCGACGTT	120
5	TTTACCATTA AATGTTATAC AGAGTAGAGT GGTAGCGACT GATATTAAAT CTATTGCTAA	180
	AGAGGCAAAC GGATTTATTA GTATCGCTTC GGAAGCAGTT AAAGTAGCAC CAGAATATCA	240
	AAATATTATC GGAATTTAT TAGGTAATAC GATTATCGTT GATCATTTAA AGCATGCAAA	300
10	TGAATTGGCA CGTGCGATTA AATATCGAAC TCGTATTGTT ACTTTGGAAG GTGATATTGT	360
	AAATCCTGGT GGtTCTATGA CTGGTGGTGG CGCTCGTAAG TCAAAAAGTA TTCTGTCTCA	420
15	AAAAGACGAG TTGACAACAA TGAGACACCA ATTAGAAGAT TACTTGCGTC AAACAGAATC	480
	ATTTGAACAA CAATTTAAAG AGTTGAAGAT AAAAAGTGAT CAATTAAGTG AACTGTATTT	540
	TGAAAAAAGT CAAAAGCATA ATACACTTAA AGAGCAAGTG CATCATTTTG AAATGGAGCT	600
20	CGATAGATTA ACTACACAAG AAACACAAAT AAAAAATGAT CATGaAGAAT TCGAATTTGA	660
	AAAAAATGAT GGTTATACGA GTGACAAAAG TCGACAAACT TTGAGTGAAA AAGAACTTA	720
	TCTAGAAAGT ATTAAAGCAT CTTTAAACG ACTAGAAGAT GAAATTGAAC GCTACACAAA	780
25	ACTTTCTAAA GAAGGTAAGG AAAGCGTTAC TAAACACAA CAAACCTTAC ATCAGAAACA	840
	ATCTGATCTT GCTGTGGTTA AAGAGCGTAT TAAACACAA CAACAGACAA TAGATCGATT	900
	AAATAATCAA AATCAACAAA CTAAACATCA ATTAAAAGAT GTTAAAGAAA AAATTGCATT	960
30	CTTTAATTCG GATGAAGTGA TGGGCGAACA AGCTTTTCAA AATATTAAAG ATCAAATTAA	1020
	TGGTCAACAA GAAACGAGAA CACGCTTATC AGATGAATTA GATAAATTGA AACAAACAG	1080
35	TATTGAGTTG AATGAACAAA TCGATGCGCA AGAAGCTAAA CTACAAGTTT GTCACCAAGA	1140
	TATTTTAGCT ATCGAAAATC ACTACCAAGA TATTAAAGCT GAACAATCAA AGCTAGATGT	1200
	ATTAATTTCAT CATGCGATAG ATCATTaAAT GATGrATATC AATTGACTGT TGAACGTGCG	1260
40	ArATCTGAAT ATACGaGTGA TGrATCGATg ACGCATTACG TAAAAAAGTT AAGTTAATGr	1320
	AGaTGyCGAT TGATGrACTA GGTCTGTAA ACTTAAATGC AATTGAACAA TTTGAAGAGT	1380
	TAAATGAACG TTATACATTT TTAAGTGAAC AACGTACAGA TCTTCGTAAA GCTAAAGAAA	1440
45	CATTAGAGCA AATTATAAGT GAAATGGATC AAGAGGTTAC TGAAAGATTT AAAGAACTT	1500
	TCCATGCTAT TCAAGGACAT TTTACAGCTG TGTTCAACA ATTGTTTGGT GGAGGCGATG	1560
	CAGAATTGCA ATTAAC TGAA GCCGATTATT TAACAGCTGG TATTGATATT GTGGtACAAC	1620
50	CACCGGGTAA AAAGTTGCAA CATTTATCGT TACTGAGTGG TGGTGAGCGT GCATTAACTG	1680
	CTATTGCTTT ACTATTTGCA ATTTTAAAG TAAGATCTGC ACCTTTTGTT ATATTAGrTG	1740
55		



	TATCAGACGA AACACAATTC ATTGTTATTA CACACCGTAA AGGAACAATG GAATTTGCAG	1860
	ATAGGTTATA CGGTGTAACA ATGCAAGAAT CAGGTGTTAC TAACTTGTG AGTGGAATT	1920
5	TAAATACAAT AGATGATGTG TTGAAGGAGG AGCAATAATG AGCTTTTTTA AACGCTTAAA	1980
	AGATAAGTTT GCAACAAATA AAGAAAATGA AGAAGTTAAA TCCTTAACAG AAGAACAAGG	2040
10	TCAAGACAAA TTAGAAGATA CACATTCTGA AGGTTCAACG CAGGACGCAA ATGATTTAGC	2100
	AGAAAATGCT GAAGTGAAAA AGAAGCCACG CAAGTTGAGT GAAGCGGATT TTGATGACGA	2160
	TGGCTTAATA TCAATTGAAG ATTTTGAAGA AATTGAAGCT CAAAAAATGG GTGCTAAATT	2220
15	TAAAGCAGGA CTCGAAAAAT CTCGTCAAAA TTTCCAAGAA CAATTAAATA ATTTGATAGC	2280
	GAGATATCGT AAAGTAGATG AAGACTTTTT TGAAGCTTTA GAAGAAATGT TAATCACTGC	2340
	AGACGTCGGT TTTAATACAG TGATGACGTT AACTGAAGAA TTACGTATGG AAGCACAACG	2400
20	ACGTAATATT CAAGATACTG AAGATTTGCG TGAAGTCATT GTTGAAAAGA TCGTAGAGAT	2460
	TTACCATCAA GAAGATkATA ATTCAGAAGC TATGAACTTA GAAGATGGTC GTTTAAATGT	2520
	CATTTTAATG GTTGGTGTGA ATGGTGTGG TAAACAACA ACAATTGGAA AATTAGCTTA	2580
25	CCGATATAAA ATGGAAGGTA AAAAAGTAAT GTTAGCTGCG GCGGATACTT TTAGAGCGGG	2640
	TGCTATTGAT CAATTGAAAG TTTGGGGCGA ACGTGTGGT GTAGACGTAA TTAGCCAAAG	2700
30	TGAAGGTTCT GATCCAGCTG CTGTTATGTA TGATGCgATT AATGCCGCTA AAAACAAAGG	2760
	TGTTGATATT TTAATCTGTG ATACCGCTGG ACGTTTACAA AATAAmACAA ATCTAATGcm	2820
	AGAATTAGAA AAAGTTAAGC GTGTAATTAA TCGAGCAGTG CCAGATGCGC CTCATGAAGC	2880
35	ATTACTATGT TTAGATGCTA CAACTGGTCA GAATGCGTTG TCACAAGCTA GAAACTTTAA	2940
	AGAAGTAACA AATGTTACAG GTATTGTATT AACGAAATTA GATGGTACAG CCAAAGGTGG	3000
	TATCGTATTA GCCATTGCTA ATGAATTGCA CATCCAGTT AAATATGTAG GTTTAGGTGA	3060
40	GCAATTAGAT GACTTACAAC CATTTAACCC TGAAAGTTAT GTCTACGGCT TATTGCTGA	3120
	TATGATTGAA CAAAATGAAG AAATAACAAC AGTTGAAAAT GATCAAATTG TAACAGAAGA	3180
	AAAGGACGAT AATCATGGGT CAAAATGATT TAGTtAAAAC GTTACGAATG AATTATTTGT	3240
45	TTGATTTTaT CAATCCTTAT TGACGAATAA ACAACGTaAT TATTTGGAAT TATTTTATCT	3300
	TGAAGATTAT TCTTTAAGTG AAATCGCAGa TACTTTTAAT GTGAGTAGaC AAGCAGTTTA	3360
50	TGATAATATA AGAAGAACTG GCGATTTAGT TGAAGATTAT GAAAAGAAAT TGGAATTATA	3420
	CCAGAAATTT GAGCAACGCC GAGAAATATA TGATGAAATG AAACCACATT TAAGTAATCC	3480
	AGAACAAATA C	3491
55		

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4253 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

10	AGTACGTTTT ATAATTATAA GTACGTAATT AACATATTAA CATATCGCAA GTATGTATTT	60
	AAATAAgATT GTTATAATTT CAAAGTTCAT CCAAGaTTAT GGCGTTTGCA TTTACCTATT	120
15	AAAAACGTTA TTATATCAAA GATGCGAAAG ATAATACGGG TTTATTTTAT GAAAGTGAGA	180
	AGGATAAAAT GGATAATGAG CAACGCTTAA AAAGAAGAGA GAATATAAGG AATTTCTCGA	240
	TTATAGCACA TATTGACCAC GGAAATCTA CATTGGCTGA TAGAATTTTA GAAAATACCA	300
20	AATCAGTTGA AACAAGAGAT ATGCAAGATC AGTTACTAGA TTCAATGGAT TTAGAAAGAG	360
	AACGTGGTAT TACAATCAAA TTAAACGCgt ACGTTTAAAG TACGAAGCTA AAGATGGAAA	420
	TACTTATACA TTCCATTTAA TCGATACGCC TGGACACGTC GATTTTACAT ATGAAGTGTC	480
25	ACGTTcTTTG GCAGCTTGTG AGGGCGCGAT TTTAGTAGTA GATGCGGCTC AAGGTATCGA	540
	AGCACAAACA TTAGCAAAATG TTTATTTAGC ATTAGATAAT GAGTTAGAGT TATTGCCTGT	600
	TATTAACAAA ATTGATTTAC CTGCTGCAGA ACCTGAACGC GTGAAACAAG AAATTGAAGA	660
30	TATGATAGGT TTAGACCAAG ACGATGTTGT TTTAGCAAGT GCTAAATCTA ACATTGGAAT	720
	TGAAGAGATA CTAGAGAAAA TAGTTGAAGT TGTGCCAGCT CCAGATGGTG ACCCAGAAGC	780
35	ACCACTAAAA GCGTTAATAT TTGATTCTGA GTATGATCCA TATAGAGGGG TAATTTCATC	840
	GATAAGAATT GTGGACGGTG TTGTTAAAGC CGGAGATAAA ATTCGAATGA TGGCCACTGG	900
	TAAAGAGTTC GAAGTAACAG AAGTTGGAAT TAATACACCT AAGCAGCTTC CAGTTGATGA	960
40	ATTAACAGTT GGTGATGTTG GTTATATTAT TGCAAGTATT AAAAATGTTG ATGATTCTAG	1020
	GGTTGGTGAC ACCATCACAT TAGCTAGTAG ACCTGCATCA GAACCATTCG AAGGTTATAA	1080
	GAAAATGAAT CCAATGGTAT ATTGCGGACT GTTCCCAATA GATAACAAAA ATTATAATGA	1140
45	TTTAAGAGAA GCATTAGAAA AATTACAATT GAATGATGCA TCATTAGAAT TTGAGCCTGA	1200
	ATCGTCACAA GCATTAGGTT TTGGTTATAG AACTGGTTTC TTAGGTATGT TACACATGGA	1260
	AATAATTCAA GAAAGAATTG AAAGAGAATT TGGTATTGAA TTAATTGCAA CTGCACCATC	1320
50	TGTAATTTAT CAATGTGTTT TAAGGGACGG TTCAGAAGTG ACGGTTGATA ACCCAGCACA	1380
	AATGCCAGAT CGTGATAAAA TTGATAAAAT ATTTGAGCCA TATGTTCTGT CAaCTATGAT	1440

	TATAAATATG GACTATTTAG ATGATATTCG TGTAATATT GTTTATGAAT TACCTTTAGC	1560
	TGAAGTTGTA TTTGATTTCT TCGATCAACT TAAATCTAAT ACTAAAGGAT ATGCATCATT	1620
5	TGATTATGAA TTCATCGAAA ATAAAGAAAG TAATTTAGTC AAGATGGATA TTTTATTAAA	1680
	TGGTGATAAA GTGGATGCGC TAAGCTTCAT AGTTCATAGA GATTTTGCAT ATGAACGTGG	1740
10	TAAAGCATTG GTTGAAAAAC TTAAACGTT AATTCCAAGA CAGCAATTTG AAGTACCTGT	1800
	ACAGGCTGCA ATAGGACAAA AAATTGTAGC GCGTACAAAT ATTAAATCAA TGGGTAAAAA	1860
	CGTTTTAGCT AAATGTTATG GCGGTGACAT AAGCCGTAAA CGTAAATTAC TTGAAAAACA	1920
15	AAAAGCAGGT AAAGCTAAGA TGAAAGCAGT TGGTAATGTT GAAATTCCAC AAGATGCTTT	1980
	CTTGGCTGTA TTGAAATGG ATGATGAATA ATTTTAAAAA ATCAATTAAC AATTTACAAT	2040
	GAATAAGTT TAATAACTAA AAAGAGGGAG CCTAGGATAA ATTAACGTCC TGGGCTTTAC	2100
20	AATGTTATAT TGGCAGCCAT CGACAGAGTT AAAATGAGCT TATAACAATG GGGCCCCAAC	2160
	ACAGAAGCTG ACGAAAAGTC AGCTTACTAT AATGTGCAAG TTGGGGTGGG GCCCCAACAT	2220
	AGAGAATTC GAAAAGAAAT TCTACAGGCA ATGCAAGTTG GGGTGGGACG ACGAAATAAA	2280
25	TTTIGCGAAA ATATCATTTC TGTCCCACTC CCTTATGCAT GAGTTTACT CATGTAATTT	2340
	TATTTTTAAG GACATATTAC ATCTGGCTAA TGTGTAAGAG CCACTACATA ATAAATCATT	2400
30	AGTGGTCTCT TATTATTTCT ATCTCACTCC CTCTAAACAA GAATAAATAT TAAATGAAT	2460
	CGATATATTA GACAATCATT GATTAAACGT TAAAGTTAAA AGTAAGAATA ATTGCAGATA	2520
	GTCCAACAGG ATATAGCCGA TTGGATAAAA AGTCTGAGAA GCGGGGCATT AAAATGACGG	2580
35	TACAAAGTGC ATATATACAT ATTCATTTT GTGTAAGAAT ATGTACATAT TGTGATTTCA	2640
	ATAAATATTT TATACAGAAT CAACCTGTAG ATGAGTACTT AGATGCACTA ATCACAGAAA	2700
	TGTCTACAGC AAAATATAGG ATCTTAAAGA CCATGTATGT AGGTGGCGGC ACACCAACGG	2760
40	CCCTTTCTAT TAATCaGTTG GAAAGATTAC TTAAAGCAAT ACGTGATACG TTTACAATCA	2820
	CAGGCGAGTA TACATTTGAA GCAAATCCTG ATGAGTTAAC TAAAGAGAAA GTCCAATAT	2880
	TAGAGAAATA TGGAGTAAAA AGGATTTCAA TGGGCGTTCA AACATTCAAG CCGGAGTTAT	2940
45	TGTCTGTTTT AGGTAGAACG CACAATACTG AAGATATTTA CACTTCGGTG TTAAATGCTA	3000
	AAAACGCAGG TATTAAATCA ATCAGTTTAG ATTTAATGTA TCATTTACCG AAACAGACGA	3060
50	TTGAAGATTT TGAACAAAGT TTAGATCTAG CTTTAGATAT GGATATTCAA CATATTTCGA	3120
	GTTACGGCTT AATACTTGAA CCTAAAACCC AATTTTATAA TATGTATAGA AAAGGCTTGC	3180
	TCAAACCTGC TAATGAGGAT TTAGGTGCTG ACATGTATCA GTTGCTGATG TCTAAGATAG	3240
55		

AACATAATAA GGTTTACTGG TTTAATGAGG AATATTATGG ATTTGGAGCA GGTGCAAGTG 3360  
 GTTATGTAGA TGGTGTGCGT TATACGAATA TCAATCCAGT GAATCATTAT ATCAAAGCTA 3420  
 5 TAAATAAAGA AAGTAAAGCA ATTTTAGTAT CAAATAAACC TTCTTTGACT GAGAGAATGG 3480  
 AAGAAGAAAT GTTCTTGGG TTGCGTTTAA ATGAAGGTGT GAGTAGTAGT AGGTTCAAAA 3540  
 10 AGAAGTTTGA CCAATCTATT GAAAGTGTCT TTGGTCAAAC AATAAATAAT TTAAGAGAGA 3600  
 AGGAATTAAT TGTAGAAAAG AACGATGTGA TTGCACTTAC AAATAGAGGG AAAGTCATAG 3660  
 GTAATGAGGT TTTTGAAGCT TTCCTAATAA ATGATTAAAA AAAATTGAAA TTTCGAGTCT 3720  
 15 TTAACATTGA CTTACTTTGA CCAATTTGAT AAATTATAAT TAGCACTTGA GATAAGTGAG 3780  
 TGCTAATGAG GTGAAAACAT GATTACAGAT AGGCAATTGA GTATATTAAA CGCAATTGTT 3840  
 GAGGATTATG TTGATTTTGG ACAACCCGTT GGTTCATAAA CACTAATTGA GCGACATAAC 3900  
 20 TTGAATGTTA GTCCTGCTAC AATTAGAAAT GAGATGAAAC AGCTTGAAGA TTAAACTAT 3960  
 ATCGAGAAGA CACATAGTTC TTCAGGGCGT TCGCCATCAC AATTAGGTTT TAGGTATTAT 4020  
 GTCAATCGTT TACTTGAACA AACATCTCAT CAAAAACAA ATAAATTAAG ACGATTAAAT 4080  
 25 CAATTGTTAG TTGAGAATCA ATATGATGTA TCATCAGCAT TGACATATT TGCAGATGAA 4140  
 TTATCAAATA TATCTCAATA TACAACTTTA GTTGTTCATC CTAATCATAA ACAAGATATT 4200  
 ATCAATAATG TACTTGTAT TCGTGCTAAT CCTAATTTAG TTATAATGGT TAT 4253  
 30

(2) INFORMATION FOR SEQ ID NO: 119:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3395 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

TCCCTAATCG AACAAAATTA TGCGCATAAA CAAAGTAGAT TGATATAAAA TTCTTAATTA 60  
 TCAGAATATA TTTACAAATC TGAATTTTAT TAGTATATTG G<sub>r</sub>TAGT<sub>r</sub>TTT ATAGAGGCAT 120  
 45 GACGGT<sub>a</sub>TTT GAGCAGGATT TTAAATCGG<sub>g</sub> ATTTTATAAT CGATTTAAGA GAGGCCAC<sub>t</sub>T 180  
 TGCTTG<sub>c</sub>ACA TTAATACTGT CAATGGGAGG GGAATGTATA TGAGT<sub>r</sub>AAGC ACATCAATTA 240  
 ATTCAAGAGG ATGAACATTA TTTTGC<sub>g</sub>AAA TCAGGACGTA TTAAATATTA TCCGTTAGTG 300  
 50 ATTGATCATG GATATGGAGC AACATTGGTT GATATTGAGG GGAAGACATA TATCGATTTG 360  
 TTATCGAGTG CGAGTTCTCA AAACGTAGGT CATGCACCTA GAGAAGTAAC AGAAGCGATA 420

55

	GTACGTTTAG CTAAGAAGCT TTGTGAGATT GCACCTGGAG ATTTTGAAAA AAGAGTGACC	540
	TTCGGATTAA CCGGATCAGA CGCAAATGAT GGCATCATTA AATTTGCCAG AGCATATACA	600
5	GGGCGTCCTT ATATCATTAG TTCTACTAAT GCATATCATG GTTCAACTTT TGGCTCATTG	660
	TCTATGTCAG CTATTAGTTT AAATATGCGC AAACATTATG GTCCGTTATT GAATGGTTTT	720
10	TATCATATTC CGTTTCCAGA TAAATATCGT GGTATGTACG AGCAGCCACA AGCTAATTCA	780
	GTAGAAGAAT ATTTAGCACC CTTAAAAGAA ATGTTTGGCA AGTATGTACC TGCTGACGAA	840
	GTAGCATGTA TTGTTATTGA AACGATACAA GGCATGCTG GACTTTTAGA ACCAGTTCCA	900
15	GGGTATTTTG AAGCGTTAGA AAAGATTTGT CGTGAACATG GTATTTTAAT CGCTGTCGAT	960
	GATATTCAAC AAGGTTTTTG GAGAACAGGT ACATGGAGTT CAGTCTCGCA TTTTAATTTT	1020
	ACGCCTGATT TAATCACTTT CGGAAAATCC TTAGCAGGTG GTATGCCTAT GTCAGCAATT	1080
20	GTTGGACGCA AAGAGATTAT GAATTGTTTA GAAGCACCAG CACATTTATT TACAACAGGT	1140
	GCTAATCCAG TTAGTTGTGA AGCTGCATTA GCCACAATTC AAATGATTGA AGATCAGTCG	1200
	CTTCTTCAGG CTAGTGCGGA AAAAGGGGAA TATGTTAGGA AACGAATGGA TCAATGGGTA	1260
25	TCTAAATACA ATAGTGTAGG CGATGTTAGA GGTAAAGGTC TGAGCATTGG TATTGATATT	1320
	GTTTCCGACA AAAAATCAA AACACGTGAT GCCAGTGCGG CACTTAAAT TTGTAATTAC	1380
	TGCTTTGAGC ATGGCGTAGT TATTATAGCT GTAGCAGGAA ATGTGTTGCG ATTCCAACCG	1440
30	CCATTGGTAA TAACATATGA GCAATTAGAC ACGGCGTTAA AACTATAGA AGATGCACTG	1500
	ACTGCTTTGG AAGCAGGTAA CTTAGATCAA TATGACATAT CTGGACAAGG TTGGTAATAG	1560
35	CGATTATCTT AATATAAAAT AAAAAATCAT TTCCACATCT GGATGTTAAT CAGATGGGAA	1620
	ATGATTTTTT TTATTTTTTA TTTGGTGCGG TGGTATTCAG CTACGTCATT TTTCTTAGAA	1680
	TGTGTAAGTC CATAACTTAA ATATAGGATG ATACCAACAA TAAACCAAAT TAAAGTGTAT	1740
40	AATTCGCTT CGAATCCTAA TCCCCAGAAT ACTAGCAATA CTAAAACAAA TGTAATTGCT	1800
	GGTAACACAG GATATAAAGG TAATTTAAAT GCAGGAATTG GTAGATCTTT ACCTTCACGC	1860
	TTTCTCAAAC GATACATTGC TAATGAAACG AACATAAATG CAACAAGTGT ACCTGCTGAA	1920
45	ATTAATTGTG CTAAAAATGC GAATGGGAAC ATAGAACCAA TTAACACACC AATAATAGTA	1980
	AGTATAACTA GTGCGCGATT AGGTAAATGT TTGTCGTTTA AGTGGCTTAA CCATGAAGGT	2040
	AATAAGCCGT CACGTCCAAA TGAATAAAGT AAACGTGAGC CTGCTAACAT CATACCAATT	2100
50	AATGCTGTAA ACATACCGAT AACAGAGATA GCTTGAACAA TAGCTGCTAC AACACCATGA	2160
	CCACTTTGAC GTAAAGCCCA ACCAACAGGT TCAGCATTGT TTGCGTATTG TGAGTAATGG	2220
55		

CCAAGAATAC CTCTAGGCAT TGTCTTTTGA GGATCAAGTG CTTCTGCTGA GTTTGCTGCG 2340  
 ATAGAATCGA AACCGATATA CGCTAAGAAA ATCATTGAAA CACCAGCATA TATGCCTTGC 2400  
 5 CATCCACCAA AGTCACCTGT AGCAGTTACT TTGTGTTCTG GAATAAATGG CACATAGTTA 2460  
 CTAACATTTA TTGCTGTTAA ACCTACGATG ACAAATAAAA TAATAGCTAA TACTTTTAAA 2520  
 ATAATAAAA TATTTTCCAT ACGAGCTGCT TCCGACATAC CACGTGATAG TAATAATGCA 2580  
 10 GTTAATAAAA TAACGATAGC AGCAATAATA TCGATAAAAC CGCCATTTGT ACCAAATGGA 2640  
 TTTGATAATG CTGCAGGTAA TTCGATGCCA ATTGGTTTCA CAAGTCCGCG TAAATTCGCT 2700  
 15 GAGAATCCTG ATGCAACAAA GGCTACGGCG ATAAATATT CAGCTAATAG AGCCCAACCG 2760  
 GCAACCCATC CAAAAAATTC ACCAAATAAT ACATTGACCC AAGAATAGGC TGAACCTGCA 2820  
 AATGGCATAG CGGCAGCCAT TTCTGCATAA GTAAATGCAA CTAAACCAGC AACAAATAGCA 2880  
 20 GCGAGTAAGA ATGATAACGC AACGGCCGGT CCTGCATGTT CTGCAGCAAC AATGCCAGGT 2940  
 AGCGTAAAGA TAGATGTCGA TACAATTGTT CCTACACCTA AAGCTAAGAA ATCACGCACC 3000  
 CGAAGTGATC GCTTTAAATG ACCATCTTTA TTTTGATAGA TAGCCGGATC CTCTTTTCGT 3060  
 25 GCTATTTTAT TGAAAAAAT TCCCATAAAC TTTCTCCCA AACATTCATA AACAAATCTA 3120  
 TACGGTGT TTAAATATGT TATATCATAG CACAAATAAT CAATATTTTG TCTAAAAATT 3180  
 CTGAAAAATC ACAACTTTAT GTTACGTATT AATGACTTGT CTTGATAACA TCCATAGATT 3240  
 30 TTTTAAATGA TAAACTGAT TATAACAGAT ATTAAATGAA TAAGTACTAT TTTTTCnAA 3300  
 TTTTCTAACA ATTTTGCACA TTATATGTTT AAAATCAATT TCATGTTTAT GGTCTGATTG 3360  
 35 GCTAGTGTGT ATGAAATGTA AnTCTTTGAC TnnGA 3395

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13508 base pairs  
 40 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

ATCAGGTAAT GCCATGCGTT TAGCTGAAAA TTTTTCAGA ACGTTTAAGT GATATCGGAC 60  
 ATCAAGTTGT TTTGATGTCA ATGGATGAAT ATGATACGAC AAACATCGCG CAGTTAGAAG 120  
 50 ATTTATTTAT TATTACGTCT ACTCATGGTG AAGGAGAACC GCCTGATAAT GCATGGGATT 180  
 TCTTTGAATT TTTAGAAGAC GATAACGCAC CTAATTTAAA TCATGTGAGA TATTCAGTAC 240

55

	TACTAGAAAA	TCTAGGCGCT	GAGCGTATAT	GTAAGCGTGT	AGATTGTGAT	ATTGATTATG	360
	AAGAAGACGC	AGAAAAGTGG	ATGGCAGACA	TCATTAATAT	TATTGATACC	ACATCAGAAG	420
5	GTATTCAAAG	TGAATCGGTG	ATAAGTGAAT	CAATTAAGTC	TGCCAAAGAA	AAGAAATATT	480
	CTAAATCAAA	TCCATACCAA	GCAGAAGTAT	TAGCGAATAT	CAATTTAAAT	GGTACCGATT	540
	CAAATAAAGA	AACACGACAT	ATAGAATTTT	TACTTGATGA	TTTTAGTGAA	TCATATGAAC	600
10	CAGGAGATTG	TATAGTAGCA	TTACCGCAAA	ACGACCCTGA	ATTGGTTGAA	AAACTAATAT	660
	CCATGTTAGG	TTGGGATCCG	CAATCTCCGG	TGCCAATTAA	TGATCATGGT	GATACAGTTC	720
	CTATTGTTGA	AGCACTAACA	TCACATTTTG	AATTTACTAA	ATTAACATTG	CCATTATTGA	780
15	AAAATGCAGA	TATCTATTTT	GACAATGAAG	AATTATCTGA	ACGTATTCAA	GATGAGTCAT	840
	GGGCGCGTGA	ATATGTTATA	AATCGGGACT	TTATAGATTT	AATAACAGAT	TTTCCAACTA	900
20	TAGAATTACA	ACCTGAGAAT	ATGTATCAAA	TCCTTAGAAA	ATTACCACCA	AGAGAGTATT	960
	CGATTTCTAG	TAGTTTTATG	GCAACGCCAG	ATGAAGTGCA	TATTACCGTT	GGTACGGTTC	1020
	GTTATCAAGC	ACATGGACGT	GAGAGAAAAG	GTGTATGCTC	GGTTCATTTT	GCTGAGCGAA	1080
25	TTAAACCAGG	CGATATAGTA	CCAATTTATT	TGAAGAAAAA	TCCGAAGTTC	AAATTTCCGA	1140
	TGAAGCAAGA	TATACCGGTT	ATTATGATTG	GACCAGGTAC	TGTAATTGCT	CCTTTTAGAG	1200
	CATATTTACA	AGAACGTGAA	GAAGTTGGTA	TGACTGGAAA	AACATGGTTG	TTCTTTGGTG	1260
30	ATCAACACCG	TAGTTCTGAC	TTTTTATATG	AAGAAGAAAT	AGAAGAATGG	CTTGAAAATG	1320
	GAAACTTAAC	ACGCGTAGAT	TTAGCATTTT	CAAGAGACCA	AGAACACAAA	GAATATGTAC	1380
35	AGCATCGTAT	AATGGAAGAA	AGTAAACGTT	TCAATGAATG	GATTGAGCAA	GGCGCACAA	1440
	CTATATTTGT	GGCGATGAAA	AATGTATGGC	GAAAGATGTC	CATCAAGCCA	TTAAAGATGT	1500
	ATTGCTAAAA	GAACGTCATA	TTTCTCAAGA	AGAAGCAGAG	TTATTATTGC	GACAAATGAA	1560
40	ACAAACAACAA	CGCTATCAAC	GTGATGTTTA	TTAGCGATTG	GTGTTAAATA	TTTAAAGGTG	1620
	TAATGATGTA	AAAAGATATA	AAGGATGTTG	CTCAACATGA	ATATGCCATT	AATGATAGAT	1680
	TTAACAATA	AAAATGTCGT	CATAGTTGGT	GGAGGCGTCG	TTGCAAGTCG	TCGGGCACAA	1740
45	ACATTAAATC	AATACGTTGA	ACATATGACG	GTCATCAGTC	CGACAATCAC	TGAAAAACTT	1800
	CAAAATATGG	TAGATAACGG	TGTCGTCATA	TGGAAAGAAA	AAGAATTTGA	ACCAAGCGAT	1860
	ATTGTAGACG	CGTATCTAGT	TATTGCAGCA	ACCAATGAGC	CACGTGTCAA	TGAAGCGGTA	1920
50	AAAAAAGCCT	TACCTGAGCA	TGCCCTTTTT	AATAATGTTG	GAGATGCATC	AAATGGCAAT	1980
	GTTGTATTTC	CAAGTGCACT	ACACCGCGAC	AAGCTAACTA	TCAGTGTATC	AACTGATGGT	2040
55							

	TACAGTTCGT ATATCGACTT TTTATATACT TGCCGACAGA AAATAAAAGT ACTTGATATA	2160
	ACATATAACG AAAAGCAACA GTTACTGTCA CAAATTGTGT CACAAGAATA TTTAAATCAT	2220
5	GACAAACAAG CTCAATTTTT AGCGTGGTTG GATGTAAGAT AATAATAGCG GACCGTCTAA	2280
	CCGTCTAAGG TAAGTCTTCT TATTTTAACT TTAACGCTTA ATCATTGAAA TTAAGACATG	2340
	GGCGGCTTTG TGAATAGTCT AATAATGAAG GATTTAAGCG ATAATGATAT GCGTTTTAAA	2400
10	TATGAATATT ACAATAGAGA AAAAGATACG TAGAACAAAC TTAATAAAAT AGGTGGATAA	2460
	ATTGAAATCT GGTGAAGTC GTTACTATCA TAGCGACCTT TAGCCAGATT TTTTGTGCAA	2520
15	TAGAAAGCAA TAATAAAAAT GATAGATCAA AATGAAATAC AGGACAGGAT ATACAAGGAT	2580
	TAGTCATGCC ATGTTATCAA GTAGGAAAAT CAAACTTCAC TATTGATAGT TACGCAAAAA	2640
	AGATTTTTTT GATAAAATGA GATAACTTAA ATATAAAAAA TTATATTAAT TATAATATTT	2700
20	AAGTTAAAGA GGGGGATTAT GTAAATTGTA TTAAAAGTGG AGGGAGAAAA TAATATGAAT	2760
	AGTGATAATA TGTGGTTAAC AGTAATGGGG CTCATTATTA TTATTTCAAT TGTAGGTTTA	2820
	CTCATTGCCA AAAAGATAAA TCCAGTTGTA GGTATGACAA TCATACCTTG CTTAGGGGCA	2880
25	ATGATTTTAG GATATAGTGT GACAGATTTG GTTGGATTTT TTGCTAAAGG GTTAGATCAA	2940
	GTCATCAACG TTGTTATTAT GTTTATCTTT GCCATTATTT TCTTTGGCAT CATGAACGAT	3000
	AGTGGTTTAT TCAAGCCGCT TGTCAAACGC TTAATATTAA TGACACGAGG CAATGTCGTC	3060
30	ATTGTCTGTG CAATGACAGC TTTAATTGGC ACAATAGCCC AATTAGATGG GGCCGGTGCG	3120
	GTAACATTTT TGCTTTCTAT TCCTGCATTA TTACCTTTAT ATAAAGCGTT AAATATGAAT	3180
35	AAATATTTAT TGATTTTACT ATTAGCATT AAGCGGCGCA TTATGAACAT GGTACCTTGG	3240
	GGAGGTCCAA TGGCTCGTGT AGCTGCAGTG TTAAAAGCCA AAAGTGTCAA TGAATTATGG	3300
	TATGGATTAA TACCTATTCA AATAATAGGT TTCATTCTTG TTATGTTGTT TGCGGTATAT	3360
40	CTTGGAATTA AAGAACAGAA ACGTATCAAA AAAGCAATAG AGAGAAATGA ATTACCGCAA	3420
	ACACAAGATA TAGATGTACA TAAATTAGTT GAAGTATATG AACGAGATCA AGATGTAAGG	3480
	TTTCCTGTAA AAGGACGTGC AAGAACAAAA TCATGGATAA AATGGGTGAA TACAGCTTTA	3540
45	ACTTTAGCTG TTATTCTATC GATGTTAATA AATATTGCGC CACCTGAATT TGCATTATG	3600
	ATAGGTGTTy CGTTGGCACT TGTTATTAAT TTTAAATCAG TGGATGAACA AATGGAACGA	3660
	TTAAGAGCgC ATGCGCCGAA TGCATTAATG ATGGCTGCAG TGATTATTGC AGCAGGTATG	3720
50	TTTTTAGGTG TACTAAATGA AACCGGTATG CTTAAAGCGA TTGCGACCAA TTTAATCAAA	3780
	GTGATTCCTG CAGAAGTAGG ACCATACTTG CATATTATTG TAGGTTTACT TGGCGTACCA	3840
55		



	ACAGCAGGGC AATTTGGTGT ACCGTCTGTA TCAACAGCTT ATTCAATGGT CATAGGGAAT	3960
	ATTATAGGTA CATTTGTCAG CCCATTTTCA CCAGCCTTAT GGTGGCAAT TGGTTTAGCA	4020
5	GAGGCAAACA TGGGCACGTA TATTAAGTAT GCATTCTTTT GGATTGGGG ATTGCTATC	4080
	GTTATGTTAG TAATTGCAAT GTTGATGGGC ATTGTGACGA TTTAAGTATG AAAAAATAGA	4140
	AACATATGGTC ACGTTGCAAA ATGAAATAAT AGTTGCATAA ACATGTCGAA ATGACGGACC	4200
10	AATCTTTAAA CAATTTTAAA AATTAATGAA ATAATTGTGT AGAAATATGA ATTTCACTAA	4260
	ATGTTAATAA CTTTGTGACG TTTTAGTTAA CAGACTAATA AAAATTTGAA AATACTATAT	4320
15	ATAGTGGTAT AACGTAATGA GTAGACACAA TATATAGGAA GAAGGGGTAA AATGAATCAA	4380
	ATCGAAGAAG CATTAAACGGG TTTGATTCTT AAAGATCCTG CTATTGTAA CGAAAATGCT	4440
	AACAAAGATA GTGATACATT TTCAACAATG AGAGATTTAA CAGCAGGTAT CGTTTCTAAA	4500
20	TCTTACGCAT TAAATCATTT ATTACCAAAG CACGTTGCAG ATGCACATCA AAGAGGGGAC	4560
	ATACATTTTC ACGACTTAGA TTATCATCCA TTCCAACCGT TAACTAACTG TTGTTTAATA	4620
	GATGCTAAAA ATATGCTACA TAATGGATTT GAAATAGGCA ACGCGAATGT AACTTCACCA	4680
25	AAATCAATAC AAACGTCATC AGCGCAGCTT GTACAAATTA TAGCCAATGT TTCTAGCAGT	4740
	CAATATGGTG GCTGTACGGT TGACCgCGTT GACGAATTAC tTAGTACATA TGCACGACcA	4800
	TAATGAAGAA CAACATAGGA ATATsCGCAA AGCAATTTGT CAAAGAATCT GAAATTGATC	4860
30	GTTATGTTGA TCAACAAGTC ACTAAAGACA TCAATGATGC GATTGAAAGT TTAGAATATG	4920
	AAATTAATAC CTTATATACA TCTAATGGAC AGACACCTTT TGTAACATTA GGATTCCGGCT	4980
35	TAGGTACAGA TCATTTAAGT CGCAAAATTC AACAAAGCTAT CTAAATACT CGTATCAAAG	5040
	GCTTAGGAAA AGACCGCACG ACAGCGATTT TCCGAAACT TGTATTTTCA ATTAAAAAG	5100
	GAACCAACTT TAGTCCGCAA GATCCGAACT ATGACATTAA ACAACTAGCA TTAAAGTGTT	5160
40	CAACGAAACG TATGTATCCA GATATTTTAA ATTATGACAA ACTCGTAGAA ATATTAGGTG	5220
	ATTTCAAAGC GCCAATGGGT TGTCGTTTCT TTTTACCAAG TTGGAAAGAT GCGGAAGGTC	5280
	ATTTTGAAAA TAATGGTCGT TGTAATCTTG GTGTGTTTAC ACTTAATTTA CCTAGAATGG	5340
45	CATTAGAATC TGCCGGTAAT ATGACGAAAT TCTGGGAAAT CTTTATGAA CGTATCGATG	5400
	TGTTACATGA TGCATTACTT TATCGTATAA ATCGTTTGAA AGATGCTGTA CCGAATAACG	5460
	CACCGATTTT ATATAAAAGT GGCGCATTTA ACTATAAATT AAAAGAAACA GATGATGTTG	5520
50	CTGAGTTATT TAAAAATAAA CGTGCAACGA TTTCAATGGG CTATATAGGG TTGTATGAAA	5580
	CAGCTACTGT TTTCTATGGT CCAGACTGGG AAACATCTCA AGAAGCAAAA GCATTTACGC	5640

55

	GGTTCAGTAT TTmCAGTACG CCGAGTGAAT CGCTAcGGAT CGTTTTTGTC GTTTAGACCA	5760
	AGAGAGATTT GGAGATATTA AAGACATTAC AGATAAAGGA TATTATCAAA ACTCTTTCCA	5820
5	TTATGATGTA CGTAAAGATG TTACACCTTT TGAAAAGTTA GATTTTGAAA AAGATTATCC	5880
	TTATTATGCG AGTGGTGGTT TCATTCACTA TTGTGAGTAT CCGAAATTGC AACACAATTT	5940
10	GAAAGCACTA GAAGCGGTAT GGGACTACTC TTATGACAAA GTTGGTTACT TAGGTACAAA	6000
	TATTCCGATT GATCATTGTT ATGAATGTGA TTACGATGGA GATTTTGAAG CAACTGAAAA	6060
	AGGATTTAAA TGCCCGAACT GTGGCAATGA TAATCCTAAA ACAGTTGATG TCGTTAAACG	6120
15	AACATGTGGT TACCTAGGCA ATCCAGTTCA ACGTCCAGTA ATTAAAGGCC GTCATAAAGA	6180
	AATTTGCGCA CGAGTAAAC ATATGAAAGC GCCTAAAGAA TGATACTTTT AGACATTAAA	6240
	CAAGGACAAG GTTATATTGC TAAATAGAA TCAAATAGCT TTGTTGACGG TGAAGGAGTA	6300
20	AGATGCACTG TTTATGTATC AGGATGTCCA TTAAATTGTG TTGGATGTTA TAACAAAGCC	6360
	TCACAAAAGT TCAGATATGG CGAGAAATAC ACTGATGAAA TATTAGCAGA AATATTAGAT	6420
	GATTGCGATC ATGATTATAT ATCTGGGCTA AGTCTATTAG GTGGCGAACC ATTTTGTAAT	6480
25	TTGGATATTA CATTAAATCT TGTCAAAGCA TTTCGAGCAC GTTTTGAAA TACAAAGACA	6540
	ATTTGGGTAT GGA CTGGATT TTTATATGAA TATTTAGCAA ATGATTGTAC AGAACGTCGA	6600
30	GAGTTATTAT CATACTTGA CGTTTTAGTA GATGGTCTAT TTATACAACA CTTATTCAAA	6660
	CCTGATTTAC CATATAAAGG TTCTTTAAAT CAACGCATTA TAGATGTACA ACAATCACTC	6720
	TCGCATGCGC GTATGATTGA ATATATAGTT AGTTGAATAT GTATTAGAAG TCAAGGTAAC	6780
35	ATTCGTTGCC TTGGCTTCTT TTTAGGTTAG GTACATAATT GAAAGTTAAT AAAAGCAATT	6840
	CTTTATAAAA ATATATTGAT AGAATATGAC CTAACAATCA TTTTGATACC AATACTAAAA	6900
	GTTGCATATC CGTTTTTTAA AAAAGTTGAA AGAGAAAAGT GGTATTTTAG TGGGAAGGAA	6960
40	GTCTAACTTT TTGGTAGCGT TTTACAATAA ATAAATATTC GTTAATAACG TATAAATATT	7020
	CTTAAATGCC ATTCTAGTAA AATTGTGTTA ATTCGTTAAA TCGTAACTTA ACACTGTTAT	7080
	TTTAGCGCTA TTAAGGTTTT GTTTATTACG GGAAAAATTA TATAAATATT CAATAATTGC	7140
45	CAAGTTTCAA ATTGATGAA ATTTGCATTA TTATTAAATG TTAGTTATTG TCAATTTTGT	7200
	GAATCAATAT AATTATTACA TTTTGAGATA AATCGAAACA GGATTCATAA AATTAATAAT	7260
	TAGGGGGAGC ACAATTGAAA AAAGAGAAAG TTATGGACTG GACGACCTTT ATAGGGACAG	7320
50	TAGCTGTACT TCTTTTTCGA GTTATACCTA TGATGGCTTT TCCAAAAGCA AGTGAAGATA	7380
	TCATCACTGG TATTAATAGT GCCATTTCTG ATTCAATTGG TTCGATATAT TTATTTATGG	7440
55		

	TTGGTAAAGC AAGTGATAAA CCAGAATTTA ATACATTTAC ATGGGCGGCA ATGCTGTTTT	7560
	GTGCAGGCAT AGGCTCTGAT ATTTTATACT GGGGCGTTAT TGAATGGGCT TTTTACTATC	7620
5	AAGTTCCACC AAATGGCGCG AAAAGTATGA GTGATGAAGC ACTCCAATAT GCGACGCAAT	7680
	ATGGTATGTT CCACTGGGGG CCAATTGCTT GGGCTATTTA TGTTCCTACCA GCATTACCAA	7740
10	TTGGTTATTT AGTATTTGTT AAAAAACAAC CGGTGTATAA AATTAGTCAA GCTTGTGCTC	7800
	CGATTTTAAA AGGTCAAACA GATAAATTTG TAGGTAAAGT TGTAGATATC TTATTTATCT	7860
	TTGATTGCT AGGTGGTGCG GCAACATCAC TAGCGTTAGG TGTGCCATTA ATTTCTGCAG	7920
15	GCATAGAAAG ATTAAGTGGT TTAGATGGTA AAAATATGAT TTTACGTTTCGCCATTTTAT	7980
	TAACAATCAC GGTATATTTT GCCATTAGTT CATATACAGG ATTGAAAAAA GGTATTCAAA	8040
	AGTTAAGTGA TATCAACGTT TGGCTATCCT TTGTACTTTT AGCCTTTATA TTTATTATTG	8100
20	GACCGACTGT TTTTATTATG GAAACGACAG TGACAGGGTT CGGAAATATG TTGAGAGATT	8160
	TCTTTCATAT GGCAACATGG TTAGAACCAT TCGGTGGTAT TAAAGGTCGA AAAGAAACGA	8220
	ATTTCCACA AGACTGGACA ATATTCTACT GGTTCATGGT GTTAGTATAT GCGCCATTTA	8280
25	TCGGTTTATT TATCGCTAGA ATTTCAAAAG GTCGACGCCT TAAAGAAGTC GTGCTAGGAA	8340
	CAATTATTTA TGAACGCTT GGATGCGTAT TATTCTTTGG TATTTTTGGT AACTATGCTG	8400
30	TGTATTTACA AATTTCTGGA CAGTTTAATG TAACACAATA TTTAAATACA CATGGTACAG	8460
	AGGCAACCAT TATTGAAGTG GTGCATCATT TACCATTCCC ATCATTGATG ATTGTACTAT	8520
	TCCTAGTATC TGCTTTCTTA TTCTTAGCAA CAACATTTGA TTCGGGTTCA TATATTTTAG	8580
35	CGGCAGCATC TCAGAAAAAA GTGGTAGGCG AACCATTACG TGCCAATCGT TTATTCTGGG	8640
	CATTTGCATT GTGCTTATTG CCATTTTCAT TGATGCTAGT TGGTGGTGAA CGTGCATTAG	8700
	AAGTATTGAA AACTGCTTCA ATACTGGCAA GTGTGCCATT AATTGTTATT TTTATTTTCA	8760
40	TGATGATATC ATTTTAAATC ATTTTAGGGC GCGATAGAAT TAAACTTGAA ACGCGTGCTG	8820
	AAAAATTAAA AGAAGTTGAA CGTCGTTTCA TGCGAATCGT TCAAGTATCa GAAGAAGAAC	8880
	AAGACGATAA TTTATAATTC AAAGCGGGTC TGGGACGACG AAATGaATTT TGTGAAAATA	8940
45	TCATTTCTGT TCCaTTCCCC TTTTTTTAGT AGCATTGTAG GATGAACTTT TAGGTTTTCA	9000
	TTAATGTTGT ACTAAAAGAT TTAATTTTTT AGTGCTCCAA GTACTTATTT ATTGTATGAA	9060
50	GCATATCTA AATCGAAGTT TGAAAGACTC TCATTGATTA TTAAATTAAA TAAAGGTAT	9120
	GCGTATGTAC AATTCAAATT AATCGAAGGA TGAAATAAAA TGAATAATCA ATTTAAAAAT	9180
	AAACAGTCCA AATTACATGA CAGTTTAGAA TCCATCACAA AAAACTTATA TGCGACACCT	9240
55		

	ACAGAATATT GTTATCTATC ATTCCGGACA CTTAGGTGAC TCCCAACAAG ACATTGCATC	9360
	ATTAGGTGGT GTTTCAAAAAG TATTGATGAA TCATGATCAT GAATCTATAG GAGGTTCTAA	9420
5	TCAAGTTGAA GCCCCTTACT TTATACATGA AAATGATGTG GCTGCACTGA AACATAAGAT	9480
	TTCTGTTCAA AAACAATTTA GTAATCGTGT AATGTTGGAT AAGGATTTAG AAGTTATTCC	9540
10	CGCGCCTGGA CATAACCAG GGACGACACT ATTTTATG GATGATGGTC ATCACCGTTA	9600
	CTTATTTACT GGAGATTTTA TATGTTTGA AGGGAAGAGA TGGCGTACAG TTATATTAGG	9660
	TTCAAGTGAT AGAGAAAAAT CTATTCAAAG TTTAGAGATG GTTAAAGAAT TAGATTTTGA	9720
15	TGTAAGTGTA CCTTGGGTTA CTATCAAAGA TGAACCGTTA GTTTATTTTG TAGAAAATGA	9780
	ATATGAAAAA CGTGAACAAA TACAAAATAT TATTGATAGA GTACGTGAGG GCGAGAATAG	9840
	CTAATTGAAA TATATTGGCG AAGCAATGTA ACGAATCTAA GAAAGCCCTA GAAAAACCT	9900
20	CCATAATTGA TTGTCATATA AAACAAAAAC GGTAATTTCT ATTTATTGAG ATAGAAATTA	9960
	CCGTTTATTT CGTGGACCTA TTGCATTGTT TTTATCATGC ATAATCATCA TTGTCGTTGT	10020
	TTGAGTCAAT TTTAATTTTC AGAATCAGAA GGCTGTTCTG GAATTGGGAA ATATTTGAAA	10080
25	ATTTACCGC TTTCAATCGC TTCGGTTAAC TGTTCCTAAC ATTCGTAATA AACATGTGTA	10140
	TGATCAAGCT GAGCTTTAAT TTTTGTGCC TCTTGTGTTT CAGCTTCAGT TAAATCACTG	10200
30	CTTTCAAGTA ATGGATTGAT AATAGCTTGA GCATCTTTTA CTGCTTCGAC ATTGATGTCA	10260
	ATTTACGCT GGAATTTTTT AGTGAAAAAG TTTTCGAAAA AGATGAAAAA GTCTTTCTCG	10320
	GCGATAAAAT GTTGTGTCG GCTTCCTCTC GTAAATGTT GTTTAACAAT ATCAAATTCC	10380
35	TGCAATTTCT TAACGCCAGC ACTCATACTT GGTTTGCTCA TTTGCAATTG ATGACGCATT	10440
	TCATCAAGCG TCATACTGCC TTCAAACACC ATTGTGCCAT ATAAGTTTCC TACACTTCTA	10500
	TTAGTGCCAT ACAAATCCAT TGTCTGTCCA ATTGAATTAA TTACAATATC TTTTGCTTGT	10560
40	TCTAATTGTT GCTGTTTGT CTGAGAACGA GTCATCATTG CACCTCCGTA CATCATTTTG	10620
	GTCACGTAA AATAAATACT AATACATTAT AAAACCTTTT CTAaaaaaAG ACATTAAAAA	10680
	TATTTAAAGC ATTAAAGTTA AATGTTTCGT TAAATAAAAA TCTAACGAAC TTACAAAAC	10740
45	TAATTCTTGA GTTGTGTTGT AAATTGACAC ATTTTTCATT TCTATGCTAA CATAAGTnTG	10800
	TAAAATTcGT TAAATAAAAA TTTAACAAAC TTAACGGrGG TTGTTGAakG GrACTTTTAA	10860
50	aACATTTATC TCAGCGTCAA TATATTGATG GTGAGTGGGT TGAAAGCGCG AATAAAAAATA	10920
	CAAGAGATAT TATCAATCCT TACAATCAAG AAGTGATATT TACGGTTTCT GAAGGGACAA	10980
	AAGAGGATGC AGAACGTGCA ATCTTAGCTG CAAGACGTGC GTTTGAGTCT GGTGAATGGT	11040

55

	AACATCgCGA	AgCgTTAGCA	CGATTAGAAA	CATTAGATAC	TGGAAAAACG	TTAGAAGAAT	11160
5	CATATGCAGA	TATGGATGAT	ATTCATAATG	TGTTTATGTA	TTTTGCTGGA	TTAGCAGATA	11220
	AAGACGGTGG	CGAAATGATT	GATTCACCAA	TTCCAGATAC	AGAAAGCAAA	ATTGTTAAAG	11280
	AACCAGTAGG	TGTAGTTACA	CAAATTACAC	CTTGGAATTA	TCCGTTATTA	CAAGCATCAT	11340
10	GGAAAAATTGC	GCCAGCGCTT	GCTACGGGTT	GTTCACTAGT	TATGAAACCA	AGTGAAATTA	11400
	CACCATTAAC	AACAATACGT	GTTTTTGAAT	TAATGGAAGA	AGTTGGTTTC	CCTAAAGGAA	11460
	CAATTAATCT	TATTCTAGGT	GCAGGTTCTG	AAGTTGGTGA	CGTAATGTCA	GGTCATAAAG	11520
15	AGGTTGACCT	TGTATCATTT	ACAGGTGGCA	TTGAGACTGG	TAAGCATATT	ATGAAAAATG	11580
	CTGCTAATAA	TGTTACGAAT	ATTGCCTTGG	AAC TTGGCGG	TAAAAATCCA	AACATTATCT	11640
	TTGATGATGC	TGATTTTGAA	TTGGCAGTAG	ACCAAGCGTT	AAATGGTGGA	TATTTCCATG	11700
20	CAGGTCAAGT	TTGTTTCAGCA	GGATCAAGAA	TATTAGTACA	AAACAGTATT	AAAGACAAAT	11760
	TTGAGCAAGC	ACTTATTGAT	CGCGTGAAAA	AAATCAAATT	AGGTAATGGT	TTTGATGCTG	11820
	ATACTGAAAT	GGGACCAGTG	ATTTCAACAG	AACATCGTAA	TAAGATCGAA	TCTTATATGG	11880
25	ATGTAGcTAA	AGCAGAAGGC	GCAACAATTG	CTGTTGGTGG	TAAACGTCCA	GATAGAGATG	11940
	ATTTAAAGA	TGGTCTATTC	TTCGAGCCAA	CAGTCATTAC	AAATTGTGAT	ACGTCAATGC	12000
30	GTATTGTACA	AGAAGAGGTT	TTCGGACCTG	TCGTTACTGT	AGAAGGCTTT	GAAACTGAAC	12060
	AAGAAGCGAT	TCAATTAGCG	AATGATTCTA	TATATGGTTT	AGCAGGTGCT	GTATTTTCTA	12120
	AAGATATTGG	AAAAGCACAA	CGCGTTGCTA	ACAAGTTGAA	ACTTGGAACG	GTGTGGATTA	12180
35	ATGATTTCCA	TCCATATTTT	GCACAAGCGC	CATGGGGTGG	ATACAAACAA	TCAGGTATCG	12240
	GTAGAGAATT	AGGCAAAGAA	GGCTTAGAAG	AGTACCTTGT	TTCAAAACAC	ATTTTAACAA	12300
	ATACAAATCC	ACAATTAGTG	AATTGGTTTA	GCAAATAAAA	ATTAGATAAG	GTGAGTGCCA	12360
40	TTGTAAGAAC	ACAAGACACT	CAC TTTGTTT	TGTATAAGTG	GCGAAATGTT	GATTGATAAT	12420
	TTGGACTAAA	CGCAAAATGA	ATCATAGATT	ATTTCAATTAC	TGTTAGTAAC	AATCGTAAAA	12480
	GGAAAAGCGA	GTGTTTTGGT	TAGCTAAGTT	TAGCAATTCA	ACGATAACCA	ATCAGCCACT	12540
45	AACAAATATT	TCATGCAATA	CTCACTTTGA	AATACAACAA	ACTTTGGAGG	TCATAACGAT	12600
	GAGTAACAAA	AACAAATCAT	ATGATTATGT	CATCATTGGA	GGAGGCAGTG	CAGGTTCTGT	12660
	ACTAGGTAAT	CGTCTGAGTG	AAGATAAAGA	TAAAGAAGTC	TTAGTATTAG	AAGCGGGTCG	12720
50	CAGTGATTAT	TTTTGGGATT	TATTTATCCA	AATGCCTGCT	GCGTTAATGT	TCCCTTCAGG	12780
	CAATAAATTT	TACGATTGGA	TTTATTCAAC	AGATGAAGAA	CCACATATGG	GCGGTCGTAA	12840
55							

TCAACGTGGT AATCCAATGG ACTATGAAGG CTGGGCAGAA CCAGAAGGTA TGGAAACTTG 12960  
 GGATTTTGCG CACTGTTTAC CGTATTTTAA AAAATTAGAA AAAACATACG GTGCAGCGCC 13020  
 5 TTATGATAAA TTTAGAGGCC ATGATGGACC AATTAAGTTA AAACGAGGGC CAGCAACGAA 13080  
 TCCTTTATTC CAGTCATTCT TTGATGCAGG TGTGAAGCA GGCTATCATA AAACACCTGA 13140  
 TGTGAATGGA TTTAGACAAG AAGGTTTTGG ACCGTTTCGAT AGTCAAGTAC ATCGTGGTCG 13200  
 10 CCGAATGTCA GCTTCAAGAG CATATTTACA TCCAGCGATG AAGCGTAAAA ACTTAACCGT 13260  
 TGAAACACGT GCCTTTGTAA CTGAAATTCA TTATGAAGGT AGAAGAGCAA CTGGTGTTAC 13320  
 15 GTATAAGAAA AATGGCAAAC TACATACCAT CGATGCTAAT GAAGTCATTT TGTCTGGTGG 13380  
 GGCATTCAAT ACGCCACAAT TACTACAATT ATCTGGTATC GGTGATTCAG AGTTCCTAAA 13440  
 ATCAAAAGGC ATTGAGCCAC GTGTTCAATTT ACCTGGTGTG GGTGAAAACCT TTGAAGATCA 13500  
 20 CTTAGAGG 13508

## (2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7646 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

GTAAGTATTG TCTTGATTTC CTAATAAAGT TATATCTTGT AATTCATCTT GTTGACGGCC 60  
 ATGTGCCATA TAAAGCGCTC CTTTAAATTT ATTTTTTTAT TATTTTGGCG TCTCGGCGTG 120  
 35 CTTTTTCAAA CATGTAATAA CTGCAACCGA TAATAACGAC GTAACCTAAT GTTGCAATAG 180  
 AATCTGGAGA TTCTCCGAAT AGAATAAATC CAAGTATTGC TGTGAAAATT ATAGATGCAT 240  
 40 ACGTAAAAAT AGAAATATCT TTTGCTGCTG CAAAACATA TGCTAAAGTA ACACCAATTT 300  
 GACCCACAGC GGCAGCTAAG CCAGCCCCTA ATAGATAAAG TATTTGCATC TGACTCATTG 360  
 GTTCATAAGT ATATGCAGTG AAAGGTATTA AAACGATGAC AGAAAATAAG GAGAAGTAAA 420  
 45 ATACTATAGT ATATGGTGCT TyTCTTGCTG TAAGTGCTCG AACACATGTA TATGCTGATG 480  
 CTGCAAAAAT ACCTGAGAAT AAGCCAGCTA ATGATGGAAT CATAGATGAT GAAAATTCAG 540  
 GTTTCACAT TAAnAGCAaC CTAAAATAGC AATTATCATT GCTGTAATTT GaTACTTCCT 600  
 50 TACCTTTTCA TGtAAGAAaa CAATGCTTaA TAAAATCGTC CAGAAAGGAT TGAGTTTCAT 660  
 TAATGAATCG GCATCACTAA GTACCATATG ATCAATGGCA TAAATATTTA ACAATACACC 720

55

	TGGCTGATGG TATTTATATA TAAAAAATAA TGGAATAAAC ATTGCTACTA AGTTTCGTGC	840
	TAATGATTTT TGAAAAACAG GAAGGTCACC TGCAAGTCTG AAAAACACTG ACATAAAACT	900
5	GAAACCAATA GCCGAAATTA AAATGGCAAT GATACCTTTT ACTTTAGGAT TCAATTTTAT	960
	CGCCTCTTTT ATATAAAATT AACGTATTTA TATTAGCATA AAACAACATG TTGTGCATAA	1020
10	ATAGTTGAAA TTTACTATAA AAAGACTATA ATAGACTGTA GCGAACAAAC GTTCTGTGTT	1080
	TATTTGTCGG AATAATAGGG CATTACACTT TTATGAATGT TTGTGTTATT ACATAAAACA	1140
	AATATCAATT CAGTATCAAG CTAATAAGCT TTTTCTTGAT TTCTGTTGAT ACAATTGAGA	1200
15	TTGACACAGA TTTAAAAAAA TCAAGTGATA TCTACTAAAA AATTTTTTTT AATTTGTTCA	1260
	AGTTTTTCTA ATTTAGTATT GGTGCCTAGT TGGAACGTTT TACGAACATT CGATTAGAAA	1320
	ATGGCACTTT AAATCATAGT GTGTCTTATG TATAATGAAA CACATAATAT AGTGTTGGTG	1380
20	AAACGAAAAA gACACAATAT CTTGTGTTTT GTATGCAAAT GCTTTATTTA TGAAGAAAT	1440
	ACATTTAAAA GTAATTAAAC ACAGAAATTT AATAGTTATT ATCAATTAAT AGTCATATTT	1500
	TTAGAAAATG TACTGAGCAA ATGGAAGATA TCCAATGATG TAAACACTAC ATATAGTGAT	1560
25	TTTTATACAT TCAACCCATA TAAGCTACTA TTTTCTCAA TATAAATCTA TGCAATTGGT	1620
	TTACATTTGA GAAAATAAGT AGCTTCATTA TAGTTAATAC AATGCTGAGA TAACCATAGT	1680
30	AACCATGTTG TTAAAGCATT TTTTAATTGG AATGACTACT TTATTTAAAA GGGTTGAAGA	1740
	AAGAAGGTGA TCCAATGAAA ATAATATATT TTTCATTTAC TGGAAATGTC CGTCGTTTTA	1800
	TTAAGAGAAC AGAACTTGAA AATACGCTTG AGATTACAGC AGAAAATTGT ATGGAACCAG	1860
35	TTCATGAACC GTTTATTATC GTTACTGGCA CTATTGGATT TGGAGAAGTA CCAGAACCCG	1920
	TTCAATCTTT TTTAGAAGTT AATCATCAAT ACATCAGAGG TGTGGCAGCT AGCGGTAATC	1980
	GAAATTGGGG ACTAAATTTT GCAAAAAGCGG GTCGCACGAT ATCAGAAGAG TATAATGTCC	2040
40	CTTTATTAAT GAAGTTTGAG TTACATGGAA AAAACAAAGA CGTTATTGAA TTTAAGAACA	2100
	AGGTGGGTAA TTTTAATGAA AACCATGGAA GAGAAAAAGT ACAATCATAT TGAATTAAAT	2160
	AATGAGGTCA CTAAACGAaG AGAAGATGGA TTCTTTAGTT TAGAAAAAGA CCAAGAAGCT	2220
45	TTAGTAGCTT ATTTAGAAGA AGTAAAAGAC AAAACAATCT TCTTCGACAC TGAAATCGAG	2280
	CGTTTACGTT ATTTAGTAGA CAACGATTTT TATTTCAATG TGTTTGATAT TTATAGTGAA	2340
50	GCGGATCTAA TTGAAATCAC TGATTATGCA AAATCAATCC CGTTTAATTT TGCAAGTTAT	2400
	ATGTCAGCTA GTAAATTTTT CAAAGATTAC GCTTTGAAAA CAAATGATAA AAGTCAATAC	2460
55	TTAGAAGACT ATAATCAACA CGTTGCCATT GTTGCTTTAT ACCTAGCAAA TGGTAATAAA	2520

	ACATTTTAA	ACGCAGGCCG	TGCCGTCGT	GGTGAGCTAG	TGTCATGTTT	CTTATTAGAA	2640
	GTGGATGACA	GCTTAAATTC	AATTAACCTT	ATTGATTCAA	CTGCAAAACA	ATTAAGTAAA	2700
5	ATTGGGGGCG	GCGTTGCAAT	TAACCTATCT	AAATTGCGTG	CACGTGGTGA	AGCAATTAAA	2760
	GGAATTAAAG	GCGTAGCGAA	AGGCGTTTTA	CCTATTGCTA	AGTCACTTGA	AGGTGGCTTT	2820
	AGCTATGCAG	ATCAACTTGG	TCAACGCCCT	GGTGCTGGTG	CTGTGTACTT	AAATATCTTC	2880
10	CATTATGATG	TAGAAGAATT	TTTAGATACT	AAAAAAGTAA	ATGCGGATGA	AGATTTACGT	2940
	TTATCTACAA	TATCAACTGG	TTTAATTGTT	CCATCTAAAT	TCTTCGATTT	AGCTAAAGAA	3000
	GGTAAGGACT	TTTATATGTT	TGCACCTCAT	ACAGTTAAAG	AAGAATATGG	TGTGACATTA	3060
15	GACGATATCG	ATTTAGAAAA	ATATTATGAT	GACATGGTTG	CAAACCCAAA	TGTTGAGAAA	3120
	AAGAAAAAGA	ATGCGCGTGA	AATGTTGAAT	TTAATTGCGC	AAACACAATT	ACAATCAGGT	3180
20	TATCCATATT	TAATGTTTAA	AGATAATGCT	AACAGAGTGC	ATCCGAATTC	AAACATTGGA	3240
	CAAATTAAAA	TGAGTAACTT	ATGTACGGAA	ATTTTCCAAC	TACAAGAAAC	TTCAATTATT	3300
	AATGACTATG	GTATTGAAGA	CGAAATTAAA	CGTGATATTT	CTTGTAACCT	GGGCTCATT	3360
25	AATATTGTTA	ATGTAATGGA	AAGCGGAAAA	TTCAGAGATT	CAGTTCACTC	TGGTATGGAC	3420
	GCATTAACCTG	TTGTGAGTGA	TGTAGCAAAT	ATTCAAAATG	CACCAGGAGT	TAGAAAAGCT	3480
	AACAGTGAAT	TACATTCAGT	TGGTCTTGGT	GTGATGAATT	TACACGGTTA	CCTAGCAAAA	3540
30	AATAAAATTG	GTTATGAGTC	AGAAGAAGCA	AAAGATTTTG	CAAATATCTT	CTTTATGATG	3600
	ATGAATTTCT	ACTCAATCGA	ACGTTCAATG	GAAATCGCTA	AAGAGCGTGG	TATCAAATAT	3660
	CAAGACTTTG	AAAAGTCTGA	TTATGCTAAT	GGCAAATATT	TCGAGTTCTA	TACAACTCAA	3720
35	GAATTTGAAC	CTCAATTCGA	AAAAGTACGT	GAATTATTCG	ATGGTATGGC	TATTCCTACT	3780
	TCTGAGGATT	GGAAGAACT	ACAACAAGAT	GTGAAACAAT	ATGGTTTATA	TCATGCATAT	3840
40	AGATTAGCAA	TTGCTCCAAC	ACAAAGTATT	TCTTATGTTT	AAAATGCAAC	AAGTTCTGTA	3900
	ATGCCAATCG	TTGACCAAAT	TGAACGTCGT	ACTTATGGTA	ATGCGGAAAC	ATTTTACCCT	3960
	ATGCCATTCT	TATCACCACA	AACAATGTGG	TACTACAAAT	CAGCATTCAA	TACTGATCAG	4020
45	ATGAAATTAA	TCGATTTAAT	TGCGACAATT	CAAACGCATA	TTGACCAAGG	TATCTCAACG	4080
	ATCCTTTATG	TTAATTCTGA	AATTTCTACA	CGTGAGTTAG	CAAGATTATA	TGTATATGCG	4140
	CACTATAAAG	GATTAAAAATC	ACTTTACTAT	ACTAGAAATA	AATTATTAAAG	TGTAGAAGAA	4200
50	TGTACAAGTT	GTTCTATCTA	ACAATTAAAT	GTTGAAAATG	ACAAACAGCT	AATCATCTGG	4260
	TCTGAATTAG	CAGATGATTA	GACTGCTATG	TCTGTATTTG	TCAATTATTG	AGTAACATTA	4320
55							



	ATGTTTTGGA GACAAAATAT ATCTCAAATG TGGGTTGAAA CAGAATTTAA AGTATCAAAA	4440
	GACATTGCAA GTTGAAGAC TTTATCTGAA GCTGAACAAG ACACATTTAA AAAAGCATTa	4500
5	GCTGGTTTAA CAGGCTTAGA TACACATCAA GCAGATGATG GCATGCCTTT AGTTATGCTA	4560
	CATACGACTG ACTTAAGGAA AAAAGCAGTT TATTCATTTA TGGCGATGAT GGAGCAAATA	4620
10	CACGCGAAAA GCTATTCACA TATTTTCACA ACACTATTAC CATCTAGTGA AaCAAACtAC	4680
	CTATTAGATG AATGGGTTTT AGAGGAACCC CATTTAAAAT ATAAATCTGA TAAAATTGTT	4740
	GCTAATTATC ACAAACTTTG GGGTAAAGAA GCTTCGATAT ACGACCAATA TATGGCCAGA	4800
15	GTTACGAGTG TATTTTtAGA AACATTCTTA TTCTTCTCAG GTTTCTATTA TCCACTATAT	4860
	CTTGCTGGTC AAGGGAAAAT GACGACATCA GGTGAAATCA TTCGTAAAAT TCTTTTAGAT	4920
	GAATCTATTC ATGGTGTATT TACCGGTTTA GATGCACAGC ATTTACGAAA TGAACtATCT	4980
20	GAAAGTGAGA AACAAAAAGC AGATCAAGAA ATGTATAAAT TGCTAAATGA CTGTATTtTA	5040
	AATGAAGAGT CATACACAAA AATGTTATAC GATGATCTTG GAATCACTGA AGATGTGCTA	5100
	AACTATGTTA AATATAATGG AAACAAAGCA CTTTCAAACt TAGGCTTTGa ACCTTATTTT	5160
25	GAGGAACGTG AATTtAACCC AATCATTGAG AATGCCTTAG ATACAACAAC TAAAAACCAT	5220
	GACTTCTTCT CAGTAAAAGG TGATGGTTAT GTATTAGCAT TAAACGTAGA AGCATTACAA	5280
	GATGATGACT TTGTATTTGA CAACAAATAA CAATTAAATt AAAAGACCTT CACATGTAAA	5340
30	GGGAAATAGC GATTCGTTTC GTCTGTCTC CTACATGTTG AAGGTCTTTT TTTATGTGTA	5400
	TCTAACTCAT TATGAGTCTG AGTAAGAAAT CAATGCTCTA AGATGTACAA TGCTATTTAT	5460
35	ATTGGCAGTA GTTGGCGGGG CCCCAACACA GAAGCAGGCG GAAAGTCAGC TAACAATATT	5520
	GTGCAAGTTG GCGGGGCCCC AACATAGAAG CAGGCGGAAA GTCAGCTAAC AATAATGTGC	5580
	AAGTtGGCGG GGCCCCAACA TAAAAGCAGG CGGAAAGTCA GCTAACAATA TTGTGCAAGT	5640
40	TCGGgCGGGG CCCCAACATA AAGAAAAACT TTTTCCTTTA GAAATTATCA CTTCCaCaTG	5700
	AGTTTTACTC ATGTATTCCT ATTTTtAAGT ACACATTAGC TGAGGCTAAT GTTAAGAACC	5760
	ACTACTTAAT CAATCATTAG TAGTTTTTAT CATTTCCACT ATTCCCaGAC ATCaAAATCT	5820
45	TAAGGTtTCT ATTTTACTTT AAGTAAACAA AATACACATT CCGAAAAATt AAATTTcAGT	5880
	TTAATTGCAA ATATCAATAA AATTGACACT AAATTATTTG AAAGGCTAIt GAAATTATGG	5940
	TCAAAAAACG CTACTATTAA TGAGAAATAT TATCAATGAT AATGATTATC ATTAATTTAA	6000
50	AGGGAGAAAA ATTTGTAAATG AAGTATTTAT TAAAGGGAAA TATTTTGCTT CTATTACTAA	6060
	TATTGTtGAC AATTATTTcG TTGTTCATAG GTGTGAGTGA ACTATCAATt AAAGATTTAC	6120
55		

	GTATTTTAAT TGCTGGAAGT TCGTTGGCTT TAGCAGGCTT GATAATGCAA CAAATGATGC	6240
	AAAATAAGTT TGTTAGTCCG ACTACAGCTG GAACGATGGA ATGGGCTAAA CTAGGTATTT	6300
5	TAATTGCTTT ATTGTTCTTT CCAACCGGTC ATATTTTATT AAAACTAGTA TTTGCTGTTA	6360
	TTTGCACTAT TTGCGGTACG TTTTATTTG TTAATATCAT TGATTTTATA AAAGTGAAAG	6420
10	ATGTCATTTT TGTACCGCTT TTAGGAATTA TGATGGGTGG GATTGTTGCA AGTTCACAAC	6480
	CTTCATCTCA TTGCGCACGA ATGCTGTTCA AAGCATTGGT AACTGGCTTA ACGGGAAGTT	6540
	TGCCATTATC ACAAGTGGAC GCTATGAAAT TTTATATTTA AGTATTCCTC TTTTAGCATT	6600
15	GACATATCTT TTTGCTAATC ATTTACAGAT TGTAGGAATG GGTAAGACT TTTACTAATAA	6660
	TTTAGGTTTG AGTTACGAAA AATTAATTAA CATCGCATTG TTTATTACTG CAACTATTAC	6720
	AGCATTGGTA GTGGTGAAGT TTGGAACATT ACCGTTCTTA GGACTAGTAA TACCAAATAT	6780
20	TATTTCAATT TATCGAGGTG ATCATTGAA AAATGCTATC CCTCATACGA TGATGTTAGG	6840
	TGCCATCTTT GTATTATTTT CTGATATAGT TGGCAGAATT GTTGTTTATC CATATGAAAT	6900
	AAATATTGGT TTAACAATAG GTGTATTTGG AACATCATT TTCCTTATCT TGCTTATGAA	6960
25	AGGTAGGAAA AATTATGCGC aACAATAATA AAAAAATAAT GCTTTTAATT GCAGTAACGT	7020
	TATTAATTAG TATGCTGTAC TTATTGTAG GTATTGATTT TGAAATATTT GAATATCAAT	7080
	TTTCAAGTCG TTTAAGAAAAG TTCATATTAA TTATTTTAGT AGGTGCTGCC ATTGCAACTT	7140
30	CAGTGGTGAT TTTTCAAGCG ATTACAAATA ACCGTCTATT GACACCATCA ATAATGGGGT	7200
	TAGATGCAGT TTATTTATTT ATCAAAGTAT TGCCAGTCTT TTTATTTGGA ATTCAATCGG	7260
35	TATGGGTTAC TAATGTATAT TTGAACTTTA TATTAACACT TATAACGATG GTGTTATTCTG	7320
	CACTAATCCT ATTCCAAGGT ATCTTTAAAA TCGGACATTT TTCAATTTAT TTTATCTTAC	7380
	TTATTTGGTGT CCTTTTAGGA ACATTTTTTA GAAGCATAAC AGGTTTTTATT CAACTGATTA	7440
40	TGGATCCTGA GTCATTTTTA GCAATACAAA GTAGTATGTT TGCTAATTTT AATGCTTCTA	7500
	ATTGCAATTT AGTTACTTTC TCAGCAGTGC TATTAGTAAT CTTATTAGTC ATTACAATTT	7560
	TACTATTGCC TTATTTAGAT GTATTGCTTT TAGGTCGTGC TGAAGCAATT AATCTTGGGA	7620
45	TATCGTATGA AAAATTAACG CGAATT	7646

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1194 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

5 ATGAATATAT TTnnAAATAA ATTATTATGG ATTGCACCAA TnGCCACTAT GATTATCTTG 60  
 GTAATCTTTT CTTTAGCTTT TTATCCTGCA TATAATCCTA AACCAAAAGA TTTACCAATT 120  
 GGTATATTAA ACGAGGATAA AGGTACAACG ATTCAAGATA AAAATGTTAA CATTGGTAAA 180  
 10 AAATTAGAGG ATAAATTATT AGATAGTGAT TCTAATAAAA TTAAATGGGT TAAGGTTGAT 240  
 AGTGAAAAAG ACCTTGAAAA AGATTTGAAA GATCAAAAAA TCTTTGGAGT AGCTATTATT 300  
 GATAAAGACT TTTCAAAGA TGCTATGAGT AAAACACAAA AAGTAGTTAT GGATAGTAAA 360  
 15 AAAGAAGAAA TGCAACAAA AGTTGCTTCA GGTGAAATTC CGCCACAAGT GGTTCACAA 420  
 ATGAAACAAA AAATGGGGAA TCAACAAGTA GAGGTTAAGC AGGCTAAATT TAAAACGATT 480  
 GTAAGTGAAG GATCAAGCTT ACAAGGTTCA CAAATTGCAT CAGCTGTGTT AACTGGTATG 540  
 20 GGTGATAATA TTAATGCTCA AATTACGAAG CAAAGTTTGG AAACATTAAAC GAGTCAAAAT 600  
 GTTAAAGTCA ATGCCGCGGA CATCAATGGT TTGACGAATC CAGTAAAAGT GGATAATGAA 660  
 AAACTTAATA AAGTTAAAGA TCACCAAGCA GGTGGTAATG CACCATTCTT AATGTTTATG 720  
 25 CCAATTTGGA TAGGTTCAAT CGTAACGTCT ATCTTATTGT TCTTTGCATT TAGAACTAGT 780  
 AACAATATCG TCGTGCAACA TCGTATCaTT GcTCAATTG GACAGATGAT ATTTGCAGTT 840  
 GTTGCGCAT TTGCAGGTAG CTTTGTTTAT ATTTATTTCA TGCAAGGCGT TCAAAGATTT 900  
 30 GATTTTGACC ATCCAAATCG TATCGCAATT TTTGTAGCAT TTGCGATTCT TGGTTTCGTG 960  
 GGCCTTATTT TAGGTGTTAT GGTATGGCTA GGTATGAAGT CAGTTCCAAT TTTCTTCATT 1020  
 TTAATGTTCT TTAGTATGCA ACTTGTAACG TTACCTAAAC AAATGTTGCC TGAAAGTTAT 1080  
 35 CAAAAATATG TATATGATTG GAATCCATTC ACACACTATG CAACAAGTGT AAGAGAcTAT 1140  
 TATACTTGAA TCATCATATT GAATTAAATA GTACAATGTG GATGTTTATA GGGT 1194

40 (2) INFORMATION FOR SEQ ID NO: 123:  
 (i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 558 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

50 GACCGACCTA TACATCCGTA TAAGTATTTT TTGATATAAG TCTTCTAAAT CATAATGATT 60  
 AAATCCAAAT GTTTTGATGC GTCGAATAAT TAATGGTTGT AGATCCATTA CTAACTTTTC 120

55

GTATTTCAAA TATTAACTA ACCCCTTCTA TCTAAAATTT AAGGTTAGTT TAATATTGTT 240  
 ACATTCAAAA TTTCAAGATG ACGGAAATGT CATTTCTTAT GATGTCCTCT TCGTATTTTT 300  
 5 TCAAATTCTG CAAGGATTTC AGAAGATAAC GGAATTCGAG TTCTTGGCTT GTTTTCACTT 360  
 ATATCATCTA ATGATTTACT CACATCAATT TCATTTTCTT TTAAATCTCT CCACATTTTCG 420  
 CGAGATGATA TTCTATATGC ACCTGATCCA AAGATAGCAT GTTGcTCACT CaTATCACTT 480  
 10 GTTACAAC TG TAATATGcTT AGtATGCTTG tCaTAAAGtT CaTAAACCAT AACGGTTCTA 540  
 ATGGAAACCA ATCAGCTG 558

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7762 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

25 GCTTCAGACA TnTGATGATA TAATCTCTCA TCATCGATTA ATTCTTTTGC AGCTTGATAC 60  
 ACATnTTGCT TATTTGTTCC AATGACTTTT AATGTGCCAG CTTCAACACC TTCAGGACGT 120  
 TCTGTAACAC TTCGCCAAA CTAAACTGG CTTATTAAAT GATGGCGCTT CTTCTGAAT 180  
 30 TCCACCTGAA TCTGTCAAAA TAAAATAAGA TTTTnTAGCA AAATTATGGA AATCTATACG 240  
 TCCAAAGGTT CAATCAATTC AATTCTGTCA TGA CTACCTA AAATCTTTTG AGCCACCTCT 300  
 CGAACTTTTCG GGTTTTATG CATTGGATAT ACCAGTGCTA AATCAGTATA CTCATCTATT 360  
 35 AAGCGTCTAA CCGCTTTAAA TATATTTTCC aTGGGTTTCC CGATATTTTC TCGTCGGTGT 420  
 GCTGTCATrA GAATGAATTT kTtGTCATGG TATTTATCCA TGATGTTAGA TTTATAATTG 480  
 TCATCAACTG TATATTTTCAT AGCATCAATC GCAGTATTAC CAGTGACAAC AACACTTTCT 540  
 40 GAATATTTCC CTTCACTTAA CAAATGCGAT GCAGCATTTT TAGTAGGTGC AAAATGTAAG 600  
 TCAGCTAATA CACCAACTAA TTGTCTATTC ACCTCTTCTG GAAAAGGTGA ATATTTATCA 660  
 45 TAACTTCTAA GCCCTGCTTC AACGTGTCCA ATCGGCACTT GGTATATAAAA TGCCGCTAAA 720  
 CCACCTGCAA ATGTCGTCAT CGTATCACCA TGTACAAGTA CCATGTCTGG TTTTCTAAT 780  
 TGAATCACTT GTTCTAATTG AGTGATTGAT TTAGAAGTTA TCTCAGAAAG TGTCTGCTCT 840  
 50 GATTTCATAA TATTCAAATC GTATTTTGGT TTGATTTCAA AGGTACTTAA TACTGAATCA 900  
 AGCATTCTC TATGCTGTGC TGTAACAACA ACAATTGGCT CGAGCATTTT TTCTTGTTCC 960

55

	ATCTTTTCA	TCAA	ACTACT	TATCTCCGAT	TCTTCTATT	AGTACCAAAC	AATCTATCTC	1080
	CAGCGTCGCC	TAAC	CCTGGT	GTGATATATG	CTTTGTCA	ATT	CAAGTGCAGC	1140
5	AATATAAATA	TCTACATCTG	GATGTGCTTC	ATGCATCTTT	TCTACGCCTT	CTGGTGCTGC		1200
	AATTAACAC	ATGAAGCGAA	TATTTTTAGC	GCCACGTTTC	TTCAATGAAG	TAATAGCTTC		1260
	AATTGCTGAT	GCGCCTGTTG	CTAACATAGG	ATCAACAACA	ATGATTTGTC	TTTCAGTAAT		1320
10	ATCTTGAGGT	AAC	TTAGCAA	AATACTCTAC	AGCCTTTAAT	GTTTCGGGAT	CTCGATATAA	1380
	ACCGATATGT	CCAACTCTGG	CTGCAGGTAC	TAAACTTAAA	ATACCATCAG	TCATACCTAA		1440
	ACCAGCTCTT	AAAATTGGAA	CGATAGCTAA	TTTTTTACCA	GCTAATCGTT	TAGCCGTCAT		1500
15	TTTAGTTACA	GCGCTTTCAA	TATCAACATC	CTGAAGCTCT	AAGTCTCTAG	TTACTTCATA		1560
	TGCCATCAAC	ATACCAACTT	CGTCTACAAG	TTCTCTAAAT	TCTTTAGTAC	CTGTATTTAC		1620
20	ATCTCTAATA	TAGCTTAGTT	TGTGTTGAAT	TAATGGATGA	TCGAAAACGT	GTACTTTACT		1680
	CATAAAAATT	ACTCCTATCT	TTGTGTATGT	TTATTGATAT	AGAGGATATT	CAGCTGTTAA		1740
	TTTCGCAACG	CGTTCTTTAG	CTTGTTGTAA	TTTTTCTTCA	TCTTTACTAT	TTTTCAATGC		1800
25	TAAACTGATG	ATTTT	TGCAA	CTTCCTCAAA	AGCTTTTTTCA	TCAAATCCAC	GCGTTGTTGC	1860
	AGCAGGTGTA	CCTAAACGTA	TACCACTCGT	TACAAAAGGT	TTTTCTTGAT	CGAACGGAAT		1920
	GGTATTTT	TG	TTACATGTGA	TACCAACTGA	ATCTAAAGTC	TCTTCAGCTT	CTTTACCAGT	1980
30	AAGTCCTATA	GACCCTTTTA	CATCAACAGC	TACTAAGTGA	TTATCTGTAC	CGCCAGAAAC		2040
	AATTCTAAAT	CCTTCATTAA	TTAATGCTTC	TGCAAGAACT	TTTGCGTTTT	TAACCACTTG		2100
	TTGTTGATAC	GTTTTGAAAT	TATTTTCTAA	CGCTTCTCCA	AAAGCAACTG	CTTTtGCTgC		2160
35	AATAACATGC	TCAAGAGGTC	CACCTTGAAT	ACCAGGGAAA	ATTGTTTTAT	CTATGTCTTT		2220
	TTTATATTCT	TCCTTACATA	AAATCATACC	ACCACGtGGT	CCGcGTAATG	TTTTGTGTGT		2280
40	TGTAGTTGTT	ACAAAATCAG	CATATTCTAC	TGGATT	TGGA	TGTAAACCTG	CCGCTACTAA	2340
	TCCTGCAATA	TGTGCCATGT	CTACCATTAA	CTTAGCGTTT	ACTTCATCTG	CGATTTCCTT		2400
	AAACTTTTTG	AAGTCAATTG	TTCTTGAATA	TGCTGATGCT	CCTGCCACAA	TAAGCTTAGG		2460
45	CTTATGCTCT	AACGCTAATT	TACGAACCTC	ATCATAATTG	ATTCGTTCTG	TGTCTTTATC		2520
	TACTCCATAT	TCAACGAAAT	TGTAGAATTT	ACCACTAAAA	TTAACAGGCG	CTCCATGTGT		2580
	CAAGTGACCA	CCATGACTCA	AATTCATACC	TAAAACTGTG	TCGCCCATT	CTAATGCAAC		2640
50	TAAGTAAACA	GCCATGTT	CG	CTTGTAACC	TGAATGTGGT	TGAACATTGA	CATGTT	2700
	TCCAAACAAT	GCTTTAGCAC	GATCAATTGC	GATGCTTTTCA	GTAACATCTA	CAA	ACTCACA	2760

55

EP 0 786 519 A2

	TTGTGCTTCC ATAACCGCTT CCGATACAAA ATTTTCCGAT GCGATTAACT CTATGTTGCT	2880
	ATTTTGTCTC TGAAATTCTC TCTCGATTGC TTCTGCGATA ACTTTATCTT GCTTGGTGAT	2940
5	ATAAGACATA AAATCTCCCC TTCTTTCAAA AAAACTTATT GGTATTTAGC ACGTTCGCCA	3000
	CCAATCTTTT TCGGCCTAGA TGTGGCAATA GTTACAATTG CCTGTCCTAC TTGCTTTACT	3060
	GAGGTCCTTA CAGGTACACA TACATGTTTA ATATGCATGC CTATTAACGT TTGACCAATA	3120
10	TCAATTCCAC AAGGAACAGT AATATGTTTG ACCACGATCG GATCCTTCAT ATGCTGAAAA	3180
	GCGTATGTTG CCAAACCTCC TCCAGCATGT ACATCTGGAA CGACGGAAAC TTCTTCCATT	3240
	GTTAATGGAT TATACTGAGA TTTTCTATT GTTATCGCTC TGTGATATG TTCACATCCT	3300
15	TGAAAAGCAA AAGTAACGCC TGTCTCTTTA CTCACAACAT CTAATGCATT AAAAATAGTT	3360
	TCTGCAACTT CCaTCGAACC GACAGTCCCT ATTTTTCGCG CAATGACTTC CGATGTTGAA	3420
	CATCCAATTA AACATATATC TCCTTTTATTA AAAAAGGACA TATCTTTTAA TTCGTCTAAT	3480
20	AACATTGTCA AATCTTTCAT AAAAGCCAC CCTTCCTAAA AATAAAAAAG GAATATAGCA	3540
	AAGTGCTACA CTCCTCTATT ATAACCTATT TAACTGTTAA CATATACTAA TTATACAGAA	3600
25	TTCTACTAG CAAATAATAT CTTTTAATTT TAAAATTAAA CTTACAAGTT CTTCATAGGT	3660
	ATGTACATAC ATTTCTTTTG TTCCACCGTA TGGATCTATA ACTTCTCCTG CTTCTTTtAC	3720
	ATATTCATGC AATGTGAAAA CATGATTTTG CAAACCAAG TGTGCCTCTA TTAATTCTTT	3780
30	GTGCGAATAC GACATCGTCA AAATAATATC TGCTTTCAAA TCTGCTTCAG TAAATTGTTG	3840
	CGATAAGGTC GTTTCAGCTA AATGATGTTT TTCAACTAAG TCTTCAACAT AATTGGAAC	3900
	ACCTTGATTG TTCACAGCGA ATATACCTCT TGATTCAAAT TGATGATTTG GCATAACCTC	3960
35	TTTGTCAATA CTTTCCGCTA ATGGGCTACG ACATGTGTTA CCTGTACAAA CGAATAAAAT	4020
	CTTCATAGTT CACATCCTTT AATAATGTGA TTACCTGCAG CTTTTAACAT GCGATTGATA	4080
	ATTGCTTCTG TATTATCATT CAGCTCAAAG CCGTATATAT ACGCCGCTGA AATATTTTCA	4140
40	TTTTCATCAA GTGAATGTAA CACATCATAA AGATTATGAC TTGCTTGTTT AACATCATTG	4200
	TCATCCTGAC ATAATTGAAT GAATTGCGCT TCACTTGGTA TAAACGCCAC CTTATTACTC	4260
	GGCACAATAA AAGCTATAGA AGACCAATCT TTACCGTCAT TTCCAATTTT GCTCTCAATA	4320
45	TCTGTAATAA TTGTAAGTGG TGTATTGGGT GAGTAATGCT TATACTTCAT ACCTGGTGCA	4380
	ATTGGCTGTT CAGTATCATT ATAATCAGCA TGGGCGATAC TATTCGGAAG TATTTCTGTA	4440
50	ATCATTGCTG CTGTTATAGA ACCAGGTCTT GCAATTTTAT AAGGAAAAGA TGTGCAATCT	4500
	AAAACCGTAC TTTCTAATCC TTCTTCACTT TGTTCAGCTT GAACAATACC ATCGATACGG	4560

55

EP 0 786 519 A2

	GCACTTGGAG CAGCTAGAGG TTCATTTATG ATTTGTAATA ATTGTCTACC TACAGAATGG	4680
	CTTGGCATTG TAACAGCAAC TGATGATAAA CCTCCAGAAA CTTTTCGACA TAGATAGCCT	4740
5	AGCTTTAACG GCAATATAAA CGAAATAGGG CCCGGCCAGA ATGCCTGCAT TAACTTTTCT	4800
	ACGCGTGGAT CCAAAGTATA TGTAATCTCT TTTAATTGAC CTTTACTGTG TATATGAACA	4860
	ATAAGCGGAT TGTGAGATGG ACGGCCTTTA GCTTCATATA TTTTAGCTAC AGCTTCTTCA	4920
10	TCTGTCGCAT TTGCTGCAAG TCCATAAACT GTTTCAGTTG GTAAACCTAT TAAACCACCG	4980
	TTTAAACAA TGTCTTTTAT TTCATTAATT TTAGGATATT GCTGTAAATC TTCATTATAT	5040
	TCTCTAACAT CCCAAATTTT AGTATCCAAC TTAATCACGC CTTTCTTATT TATCATAATA	5100
15	TAAAGCAAAA AGCTATGCAC TTAATAATC ATAGCAAAGG CATAACTTCT AATTACCATT	5160
	TAAATGAGAC GATTTCGATCG TGGCCATTTA TATCTTTAAT AATGTCGATT TTTTGTGACG	5220
	GAAATTTTAT TAAATTTATT GATTTAAGTG CCTCACCTTG ATTGTAACCA ATTTCAAAAA	5280
20	CAACTGGGCT GCCTTTTTCC ATAACGTGAG GTAAATCTTC AATGATTGAT TCATAAATAG	5340
	CATATCCATG GTTATCTGCA AACAATGCCT GATGTGGTTC GAATCTCGTA ACCGTTGGAG	5400
25	ACATCGTAAC CATATCTTTT TCATCTATAT ATGGTGGATT AGATATCAAG CCGTTCAACT	5460
	TGATACCTTC ATTAATTAAG GGCTTTAATG CATCCCCTGT TAAAAATTGT ATTTGTGATT	5520
	GATGCTTCTC AGCATTATTA CGAGCCATAT TCATTGCTTC AAGTGAAATA TCAGTAGCAA	5580
30	TAACATTTAA ATCCGGCTTT TCACATTTCA AAGTAATTGC AAGTACACCA CTACCCGTTT	5640
	CGATATCTAC GATTGTTGCA TCATCTTCTA ACTGTTGTAA GAAATGCAAC ATTACTTCTT	5700
	CAGTTTCAGG TCTTGGTATC AAACAATTTG AGTTTACATC AAACGTTCTA CCATAAAATG	5760
35	AGGCAAAGCC AACTATATAC TGTATAGGCT CTCCTAATAA CATACGTTGT AATGCTAAGT	5820
	CGAATCTCAT AATCATCGCT TTCGGCATAT CATCATGCAT GTGGACTACA AAGTCCGTAC	5880
	GCGTCCATTG AAATACATCT AACATTAACC ATTCAGCTCG TGTGTGTTCA AACCTTTTTT	5940
40	GTTGTGTTAA ATGAATTGCT TCATCTAACT TTTCTTTATA ATTCACCATT ATTAAGTTCT	6000
	TTCAATTTAT CTGTCTGCTC TGATAAAGTC AGTGCATCTA TAATTTCTTC TAAATGGCCT	6060
45	TCCATAATTT GCCCTAATTT TTGAAGCGTT AGACCTATAC GATGGTCTGT TACACGGCTT	6120
	TGTGGATAAT TATAAGTTCG AATACGTTCT GAACGATCAC CAGTACCGAC TGCTGATTTA	6180
	CGTTGTGACG CATACTTTTG TTGTTCTTCT TGAACCTTCA TATCGTATAA ACGTGCTTTT	6240
50	AACACTTTCA TTGCTTTTTT ACGGTTTTGA ATTTGAGACT TCTCAGAAGA TGTTGCAATG	6300
	ACACCAGTTG GTAAATGGGT AATACGTACT GCAGAGTCAG TTGTGTTTAC GTGCTGACCA	6360

55

	ACATCTTCAA CTTCTGGTAA AACTGCCACT GTAGCTGTTG AAGTATGAAT ACGTCCACCT	6480
	GATTCTGTTT CAGGCACACG TTGAACGCGG TGCGCACCAT TTTCAAATTT CAATTTACTA	6540
5	TACGCGCCAT TACCAGAAAC TGAGAAACTA ATTTCTTTGT AACCACCATG GTCACCTTCA	6600
	GACGCTTCTA CTATTTCACT TTTGAATCCT TGTGATTCAG CATACTTTGA ATACATACGC	6660
	ATTAAATCAC CAGCAAAAAT CGCAGCCTCA TCACCACCTG CTGCTGCTCT TATTTCTACA	6720
10	ATAACGTCTT TGTCACTATT AGGATCTTTA GGAATCAATA ATATTTTAAG CTCTTCTTCA	6780
	AGATTTGGAA GTTCAGCTTT AATACCATTA CTCTCCTCTT TTAACATTTT TACTTCTTCT	6840
	TTATCATCAG TCTCACTTAA CATTTCTTCA ATATCAGCTA ATTCTTCTTT TTTAGCTTTA	6900
15	TAGTTACGAT AAACATCTAC AGTTTTTTGT AAATCAGCTT GCTCTTTAGA ATATTTACGT	6960
	AATTTATCTG AATCATTAC AACATCTGGG TCACTTAACA GTTCATTTAA CTGTTTCGTAT	7020
20	CTTTCTTCTA CAATATCTAA TTGATCAAAC ACTTATAATT CCTCCTTATT ATTATCACTA	7080
	GGTGCTACGA TATGGTGCGC GCGACAACGT GGCTCATAAC TTTCATTGGC ACCTACTAAG	7140
	ATAATCGGAT CATCGATTTT AGCTGGTTTA CCATTTATTA ATCGTTGCGT TCTACTAGAT	7200
25	GAAGAACCAC AAACAGCACA AACTGCTTGA AGTTTCGTTA CTTGTTCACT GACAGCCATC	7260
	AATTTAGGCA TTGGTTCGAA CGGTTGCCCC CTAAAATCCA TATCTAATCC AGCAACAATA	7320
	ACACGGTGTC CATCTGCTGA TAGTTTTTCT ACTATACTTA CAATTTCACT GTCAAAAAAT	7380
30	TGCACTTCGT CTATTCCTAT AACATCAACA TTAGTTAAGT CGTGCGTCAT AATTTCACTT	7440
	GCTTTAGAAA TATTAATCGC TTCAATGGCA TTACCATTAT GAGAGACCAC TTTTCTTTA	7500
	TGATATCGAT CATCAATCGC CGGTTTAAAT ACAACGACTT TTTGTTTAGC GTATATACCC	7560
35	CTTCTTAGAC GTCTTATTAG TTCTTCGGAT TTACCGCTAA ACATACTACC TGTAATACAT	7620
	TCTATCCAAC CGGAATGGTA AGTTTCATAC ATTGAGAGTA CCACCTTTTT CAAAACATAA	7680
40	TCGCTTTATT ATATCATATT TCAAATATTC ATAAATGTCT TTnTCATAAT TATATCGATA	7740
	TTGTACATGA ACAATTATTT TA	7762

(2) INFORMATION FOR SEQ ID NO: 125:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2583 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:



	TAAAAAATT ATTATCAATG ATGAACTAGA ATTGACTGAA TTCCACCAAG AACTTACTTA	120
	TATTTTAGAC AACATAnAAG GGAATAATAA TTATGGTAAG GAATTTGTTG CAACCGTTGA	180
5	AGAAACATTC GACATTGAAT AaAGCGGGT GgaAGCACTA TGAATCAATG GGATCAGTTC	240
	TTAACACCTT ATAAGCAAGC GGTTGATGAG TTGAAAGkGA AcTTaAAGGC ATGCGCAAAC	300
	AATATGAAGT TGGTGAACAA GCGTCGCCAA TAGAATTTGT TACTGGTCGT GTTAAACCAA	360
10	TCGCTAGTAT TATAGATAAG GCAAACAAAC GACAAATACC ATTTGATAGG TTAAGAGAAG	420
	AAATGTACGA TATCGCTGGT TTAAGAATGA TGTGCCAATT TGTTGAAGAT ATTGATGTTG	480
	TCGTCAATAT TTTAAGACAA AGAmAAGATT TTAAAGTAAT TGAAGAACGA GATTATATTC	540
15	GTAACACTAA AGAAAGTGGT TACCGCTCGT ATCATGTCAT TATTGAATAT CCAATTGAAA	600
	CATTACAAGG CCAAAAATTT ATATTGGCTG AGATTCAGAT TCGTACATTA GCAATGAATT	660
20	TCTGGGCAAC GATTGAACAT ACTTTACGAT ATAAATATGA TGGTGCTTAT CCGGATGAAA	720
	TTCAACATCG TTTGGAAAGA GCGGCAGAAG CAGCGTATTT ACTTGATGAA GAGATGTCTG	780
	AAATTAAAGA TGAAATTCAG GAAGCTCAAA AATATTACAC GCAAAAACGT TCTAAAAAC	840
25	ATGAAAATGA TTAACGAGGT GTTATAAATC ATGCGTTATA CAATTTTAAC TAAAGGTGAC	900
	TCCAAGTCTA ATGCCTTAAA GCATAAAATG ATGAACTATA TGAAAGrTTT TcGCATGaTT	960
	GaGGATrGTG AAAaTCCTGA AATTGTTATT yCAGTTGGTG GTGACGGTAC ATTACTACAA	1020
30	GCATTCCATC AGTATAGCCA CATGTTATCA AAAGTGGCAT TTGTTGGAGT TCATACAGGT	1080
	CATTTAGGAT TTTATGCGGA TTGGTTACCT CATGAAGTTG AAAAATTAAT CATCGAAATT	1140
	AATAATTCAG AGTTTCAGGT CATTGAATAT CCATTGCTTG AAATTATTAT GAGATACAA	1200
35	GACAACGGCT ATGAAACAAG GTATTTAGCA TTAAATGAAG CAACGATGAA AACTGAAAAT	1260
	GGCTfCAACAC TTGTTGTGGA TGTTAACCTA AGAGGGAAAC ACTTTGAGCG ATTTAGAGGC	1320
40	GATGGATTAT GTGTATCAAC ACCTTCGGGT TCAACGGCTT ATAACAAAGC GCTAGGTGGC	1380
	GCACTGATAC ATCCTTCACT TGAAGCAATG CAAATTACAG AAATTGCCTC GATAAATAAT	1440
	CGTGTGTTTA GAACGGTAGG ATCACCACCT GTATTACCAA AGCATCATAC ATGTTTAATA	1500
45	TCACCAGTTA ATCATGATAC CATTAGAATG ACGATAGATC ATGTTAGTAT CAAACATAAA	1560
	AATGTTAATT CAATACAATA CCGTGTAGCA AATGAAAAAG TGAGGTTTGC ACGTTTTAGA	1620
	CCATTCCCAT TCTGGAAACG TGTGCACGAT TCTTTCATAT CAAGTGATGA AGAACGATGA	1680
50	AATTTAAGTA TCATATATCA CAACAAGAAA CTGTAAAAC TTTTTTAGCA CGACATGATT	1740
	TTTCTAAGAA GACAGTGAGC GCCATTAAAA ATAATGGCGC TTTAATTGTT AATGATGAAC	1800

55

AAATACCGAG TGTTAATTTA ATACCTTATG CTCGTAAGCT AGAAGTATTG TATGAAGATG 1920  
 CTTTTATCAT CATAGTTACT AAACCAAACA ATCAAAATTG TACGCCTTCG AGAGAACATC 1980  
 5 CTCATGAAAG TTTAATCGAA CAAGTACTAT ATCATTGTCA GGAACATGGT GAAAATATTA 2040  
 ACCCACATAT TGTTACGCGT CTAGATCGTA ATACAACTGG TATTGTGATA TTCGCTAAAT 2100  
 ATGGACATAT CCATCATTTA TTTTCTAAAG TAAACTTGAA AAAAATATAT ACTTGCCTTG 2160  
 10 TATATGGTAA AACCCATACA TCTGGTATTA TTGAAGCTAA TATTAGACGG TCAAAGGATA 2220  
 GGATTATAAC TAGAGAAGTT GCCTCGGATG GTAAATACGC TAAACATCT TATGAAGTAA 2280  
 TAAATCAGAA TGATAAATAC AGTTTATGCA AAGTTCATTT GCATACGGGA CGTACACATC 2340  
 15 AAATTCGTGT ACATTTTCAA CATATTGGGC ATCCAATTGT GGGAGATTCT TTGTATGATG 2400  
 GTTTTCATGA CAAAATTCAT GGTCAAGTAC TGCAATGTAC GCAAATATAT TTTGTTTCATC 2460  
 CAATCAATAA GAACAATATT TATATTACAA TTGATTATAA GCAATTACTT AAATTATnCA 2520  
 20 ATCAACTCTA ATnCACACAG GGGGTGTAAG TATGTCAATG AnCACAGATG AAAAAGAGCG 2580  
 TGT 2583

25 (2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1818 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

35 ATCAAGTGAT ACATTTAACT GGTAAAGGAT TAAnAGATGC TCAAGTTAAA AAATCnGGAT 60  
 ATATACAATA TGAATTTGTT AAAGAGGATT TnACAGATTT ATTnGCAATT ACGGATACAG 120  
 TAATAAGTAG AGCTGGATCA AATGCGATT ATGAGTTCTT AACATTACGT ATACCAATGT 180  
 40 TATTAGTACC ATTAGGTTTA GATCAATCCC GAGGCGACCA AATTGACAAT GCAAATCATT 240  
 TTGCTGATAA AGGATATGCT AAAGCGATTG ATGAAGAACA ATTAACAGCA CAAATTTTAT 300  
 TACAAGAACT AAATGAAATG GAACAGGAAA GAACTCGAAT TATCAATAAT ATGAAATCGT 360  
 45 ATGAACAAAG TTATACGAAA GAAGCTTTAT TTGATAAGAT GATTAAAGAC GCATTGAATT 420  
 AATGGGGGGT AATGCTTTAT GAGTCAATGG AAACGTATCT CTTTGCTCAT CGTTTTTACA 480  
 50 TTGGTTTTTG GAATTATCGC GTTTTTCCAC GAATCAAGAC TTGGGAAATG GATTGATAAT 540  
 GAAGTTTATG AGTTTGTATA TTCATCAGAG AGCTTTATTA CGACATCTAT CATGCTTGGG 600

55

CTCATGTTAA AGCGCCACAA AATTGAAGCA TTATTTTTTG CATTACAAT GGCATTATCT 720  
 GGAATTTTGA ATCCAGCATT AAAAAATATA TTCGATAGAG AAAGACCTAC ATTGCTGCGT 780  
 5 TTAATTGATA TAACAGGATT TAGTTTTCTT AGCGGTCATG CTATGGGATC AACTGCATAT 840  
 TTTGGAAGTG GTATCTATCT ATTAAATCGA TTAAATCAAG GTAATTCAAA AGGTATTCTT 900  
 ATAGGGTTAT GTGCAGCTAT GATTTTATTG ATTTCCATAT CACGTGTATA TCTAGGTGTA 960  
 10 CATTATCCAA CAGATATTAT TGCCGGCATT ATTGGTGGAT TATTTTGcAT TATTTTATCA 1020  
 ACGTTATTAC TTAGAAATAA ATTAATAAAT TAAATAGTAA AAAACAAAA GCAGTAAACC 1080  
 TAAAGTGTG TAAGGGTTTA CTGCTTTTAT AAAACGTTGT TATAACGTAT ATTGTCTTTT 1140  
 15 ACGGGCATAT AAnAGGGGAA TATTTGAnAA TGACCAATCC AACAAGAACG AACGTTGTG 1200  
 GGGGGGATGT TCTATGTGGT ATTGATAATC ATTTTCAACT ACTATTATAC ATTAGTGAGA 1260  
 ATCATTGTCA ATTAGAACT AAACTTTTT TTGAATATTT TTTAAGAATA GTAAATAAAA 1320  
 20 CGCATGATTA CGCTATTTTA GAAATAAAAA AAATTTGTAT TTCTCATTAG AATTAGAATA 1380  
 TTTAAAAGTG ATGAGGTTTA AACATTATAT TGTTTACATA CTCCTTTTGA ATTCATACAT 1440  
 TATGAAATGT tACTTCCAAG TTCAAAATCG CACATTGAAA TGATGTGTGA AATGTTTAAA 1500  
 25 CTACGGTCAT tTTGTGmAAA TAAAGrTAAT AACTATTCAT TTTACAATAG TGAAAAGTCA 1560  
 GTATATGACA ACAATTAATA TTGCGGTAAG GCCTTGTTT ACAGTATTCT ATATTTAAGT 1620  
 30 ACTGCAATCA GAATTAACAG AATGCCATTA ACTGATTATT AAATATTTGA GTTAATAAAT 1680  
 AATTAATGAT TGTAGCTTGA AAAATTTAAA ACATGGTTAT TGATTTGTGA TAAAATTTAA 1740  
 ACGTAAACAA ACTAATTTAA AAAGCAACTA TTGTATAGAA AAATACAAAA TTTAAATAT 1800  
 35 ATTACCTTAT TAGAAAAA 1818

## (2) INFORMATION FOR SEQ ID NO: 127:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 12658 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

TGTTTAAACA ATAGGGGGAA TCTTATGATT GAAAAATTAG TAACCTTTTT AAATGAGGTT 60  
 50 GTTTGGAGTA AGCCATTAGT TTATGGTTTG CTAATTACTG GTGTGCTATT TACATTGCGT 120  
 ATgCGATTTT TTCAAGTTAG ACATTTTAAA GAAATGATTC GATTAATGTT TCAAGGAGAG 180

55

	GGTACAGGTA ATATTGTCGG TGTATCTACT GCAATATTTA TAGGAGGACC TGGTGCAGTA	300
	TTTTGGATGT GGATTACTGC GTTTTtagGT GCAAGTAGTG CTTTtATTGA ATCTACACTT	360
5	GGTCAAATAT TCAAGAGAGT TGAAAATAAT GAATACCGTG GTGGACCAGC GTATTATATT	420
	GAATATGGTA TTGGTGGTAA ATTTGGTAAA ATTTACGGAA TTATCTTTGC TATTGTTACG	480
	ATTATCTCAG TAGGTCTATT GCTTCCTGGT GTGCAATCTA ACGCTATAGC AAGTTCTATG	540
10	CATAATGCGA TTCATGTTCC ACAATGGTTA ATGGGTGGTA TTGTTGTAGT TATTTTGGGA	600
	TTAATTATTT TTGGTGGTGT ACGTATTATT GCCAATGTTG CAACAGCCGT TGTACCATTt	660
	ATGGCAATTA TTTACATACT GATGGCTGTC ATTATCATTT GTATCAATAT ACAAGAAGTG	720
15	CCAGCGTTAT TTGCATTAAT TTTCAAATCA GCATTTGGAT TACAATCTGC TTTTGGTGGT	780
	ATCGTTGGCG CAATGATAGA GATTGGTGTt AAACGTGGAT TATATTCAAA TGAGGCTGGT	840
20	CAAGGTACAG GTCCACACGC AGCAGCGGCa gcAGaAGTAT CACATCCAAG TAAACAAGGT	900
	CTAGTACAAG CATTttCAGT TTATATTGAT ACATTATTTG TATGTACTGC AACTGCTCTG	960
	ATTATACTTA TTTCTGGTAC ATATAATGTG ACTGATGGTA CGGTtAATGC GAATGGCACA	1020
25	CCGCATTtAA TTAAAGATGG CGGTATTTAT GTTgAAAATG CAACAGGTAA AGATTATTCA	1080
	GGTACTGCGA TGTATGCACA AGCCGGCAt GATAAAGCGT TCCATGGCAG TGGTTATCAA	1140
	TTTGATCCTA CTTTCTCTGG CGTAGgTTCG TACTTTATTG cATTtGCTTT ATTCTTCTTT	1200
30	GCATTtACTA CAATTTTGTC GTACTACTAC ATTACAGAAA CAAATGTTGC TTATTTAACG	1260
	CGTAATCAAA ATAATCAAGT TTCATCGATA TTTATTAATA TTGCTCGTGT GATTATTTTG	1320
	TTCGCTACAT TTTACGGTGC AGTTAAACA GCTGATGTAG CATGGGCATT CGGTGATTtA	1380
35	GGTGTAGGTC TAATGGCTTG GTTAAATATC ATTGCGATTT GGATTTTACA TAAGCCTGCC	1440
	GTAATGCTT TAAAAGATTA TGAAATTCAA AAGAAACGTT TAGGCAACGG TTATAATGCA	1500
40	GTTTATCAAC CTGATCCGAA TAAATTACCT AATGCTGTCT TTTGGTTGAA GACGTATCCA	1560
	GAACGTTtAA AACAAGCACG TGCCAAAAAG TAATCTACTT TTGTTTATAG TATATGTAGT	1620
	GATCATTTGA TAAAAAGAA AAGTATTGAG AATTTTAGGt GCTCAGAAAT TTGAATTTTA	1680
45	AAAATATAGT GTCTCTTGGT ACAATAACAA TACAATACT AGGGGCACTT TTTTATGTCA	1740
	GAATTTAAAA CTGGTAAGAT TAATAAACAT GTTTTATATA GTAATATTTT AAATAGAGAT	1800
	GTCACGTtAA GTATTTATTT ACCAGAATCT TATAATCAAC TTGTtAAATA TAATGTCATT	1860
50	CTTTGCTTTG ACGGATTAGA TTTTTACGT TTCGGGAGAA TACAACGTAC ATATGAATCG	1920
	TTAATCAAAG AAGCGCGTAT TGATGATGCG ATCATtGTTG GATTCCATTA TGAAGACGTT	1980

55

	GTCGGTAAAG AAATATTGCC ATTTATTGAC TCGACGTTTT CTACACTGAA AGTAGGTAAT	2100
	GCAAGGTTAT TAGTAGGGGA TAGTTTAGCG GGTAGTATTG CCTTATTAAC GGCGTTGACC	2160
5	TATCCAACGA TTTTGTAGTCG TGTAGCAATG TTAAGTCCAC ATTCAGATGA AAAAGTATTA	2220
	GATAAGCTAA ATCAATGTGC AAATAAGAA CAATTGACAA TTTGGCATGT CATTGGTCTA	2280
	GATGAAAAAG ATTTTACTTT ACCAACAAAT GGTAAAGCGTG CCGATTCTT AACACCGAAT	2340
10	AGAGAATTAG CTGAACAAAT TAAGAAATAT AATATAACTT ATTATTACGA TGAATTTGAT	2400
	GGTGGTCACC AATGGAAAGA TTGGAACCA TTGCTGTCAG ATATAATTATT GTATTTTTTA	2460
	AGTAAAAACA CAGATGATCA ACTTTATGAA TAATTTACAT TAGTAGATTT AGTATGAATT	2520
15	GTCTTCATAT AGTCTGGTCT ATAATATAAT TTATAAAGA TTTTACTGTT TAATTTAATT	2580
	TAAATTTGAC GAAATTGCAA AAGATGTATA ATGAATTATT TTTAATGTAA CGGTTTTCAA	2640
20	AGAAATTTGA TATAATAGCA ATAGGTTAAA CAAAGGAGGA ATTCAGATGA TTTTAGGATT	2700
	AGCATTAAAT CCATCAAAGT CATTTCAAGA AGCGGTGGAT TCTTACCGTA AAAGATATGA	2760
	TAAACAGTAT TCACGAATTA AACCACATGT GACAATTAAA GCGCCATTTG AAATTAAAGA	2820
25	TGGTGATTTA GATTCTGTCA TTGAACAGT TAGAGCTCGT ATTAATGGTA TACCAGCAGT	2880
	AGAAGTTCAT GCTACAAAAG CTTCTAGCTT CAAACCAACG AACAATGTGA TTTACTTTAA	2940
	AGTTGCGAAG ACGGACGACT TAGAAGAATT GTTTAATCGC TTTAATGGAG AAGATTTCTA	3000
30	TGGAGAAGCT GAACATGTTT TTGTGCCACA CTTTACAATA GCACAAGGAC TATCTAGCCA	3060
	AGAATTCGAA GATATTTTTG GTCaAGTAGC ATTAGCTGGG GTAGACCATA AAGAAATTAT	3120
	CGATGAATTA ACTTTGTTAC GTTTTGACGA TGACGAAGAT AAATGGAAAG TTATTGAAAC	3180
35	GTITAAATTA GCTTAAGTAA CATAATAGTA TTGTTAATCG TAGTATGTTT GAATTAATAA	3240
	GAAAATGGTC ATTTTTATTG AATGTAATAA AAATGACCAT TTTCTTTATT TTAAAATACG	3300
	TTTTAACCTT ACTTAGCTTT TTCTCTATTT ACTATAAAGT TGGCTTCATA AAATACAGCT	3360
40	AAGACTAAAA AGATTAATGC CGAGAAATAA AATGTATTGT TTAAATTGTT GGTAAATTGT	3420
	GTAATTAATC CGCCAAATAA TGGCCCTATC ATTGAGCCGA ATCCTTGGAT ACTATTAAAA	3480
45	ACACCCCAAG TTTCTTCTTG TTCATCTGAT TTGATAAATC GTGCCATAAA GGTATTCCAT	3540
	GCTGGTAATA AGATGCCATA CATTAGACCG ATAGCTAAAG CGATAATCCA CAAGATGTGA	3600
	ATATTAACAA TCATAGATAG AGTAAAAATT AATATCATGT ATAAAATAAA TCCGCTTAGA	3660
50	ATAACACCAT ACATAAAGTT TCTGCTGCGG TTATCTATTA GTTTCGATAA AAATAGCATC	3720
	GAAACTGCAC AGCCGATACC ACCAATAATG ATTGCAACAG TATATTCAAT TGTGCTTACG	3780

55

	TGTAAAAGAA TACCAGGGAA CaACAATAAA TGGcGCTTTG TCACATCAAC AATTTGTCTC	3900
	AATTGAGCTT TAACTGGACG AGTATTATAA TTTGTAACT TTACATCGAC AAAATAATAT	3960
5	AATATCCATG CAATTAAAAC GACTAAAGAC ATCATGAAGG CAAAGCGTGT TGGGTGCACT	4020
	TTGATAAGTA GATTCATAAA AACCATACCT ACCAATAGGC CTAACAACCA TGAAAAATAA	4080
	ACATAGCCCA TTTGTTTGCC ACGTTTATCT TCTTCAACAC TGGATAACAT AATGACCCAA	4140
10	ATAGGACTAA CTGCAATACC GAGCATCATA GCACTAAATA TGATTACAAA AGGTGATGCT	4200
	GGAAACCAAA TAACTAAAAA TAACTTGTA AATGCTAAAA TAAATCCAGT CGTTAAACG	4260
15	ATTTTGTGTC CGAATTTTTT CAGTAAAAAT CCTATAACAA AGTTTGTAGA TGCATCAGCA	4320
	ATAAAATGTA TTGAAAATGC TAGAGACGTT ATTGCTACAG CAATGGATGT AACTGTTGGC	4380
	AAGAAATTAA TATAGCTTAG GATATACATG CCTCTCGCAA ATTCCATTAA AAATAAGATA	4440
20	ATAAGCaTTA AAATGAAATT TTTATGATTA GCGTAATTAT TTAACGAAGA ATCTTGATA	4500
	TAAAGGAACC TTTCCATAAA TCTCTTGTTG TTGTGATGAA TGACCGATTA AATCAAGTAA	4560
	GTCTCGACAT ATTGTCTGTG TAGCATACTT AATTTTATCT TGTTCATTG TACTAATCAT	4620
25	GTTAGTTAAT TGCTCATTAC CGTTAGTTAA ACTTGCTACA ATTTTATTG CTTCTTCTGG	4680
	AGTATCAGCG ATTTTACCAA AACCTTTTTC TTCAAAGTAA AGGGCATTIT CAAGCTCTTG	4740
	ACCAGGTGCA GGATTTAGGA AAATCATTGG AATACAACGG GCGAAACCTT CAGTTATTGT	4800
30	GATACCACCA GGTTCGTAA TCATAAGTTG ACTTGATGCC ATCCATTCTAT TCATGTGTTT	4860
	GGTATAACCT AGAATCAATA CATTCTCGTT AGATTTAAAC TTAGCTGTTA AAGAACGCTT	4920
	TAGCTCTTTG CTCTTACCAC AAATCATAAC TACTTGTCa TTTGCaCTTT tCGCTAATAT	4980
35	ATCAGTAATC ATCGTGTCa AACCTTTAGA TACACCAAAT GCACCAGCTG aCATTAAAAT	5040
	AGTTTGCTTA TCTGGATCTA AGTTGTTGTC TATTAACCAC TGCTTTTGAT TAATAGGCGT	5100
	TTCAAATTTG TTATCAATAG GAATACCTGT CaCTTTAACT GTTGAAGGAT CAATACCTAC	5160
40	GTCTATGAAG TCTTGTTTCG TTTCTTTTGT TGCCACATAA TATCTTGTTG AATACGGCGT	5220
	AATCCAGTTT TTATGTAAGC GATAGTCTGT CATCACTGTA GCAACTGGAA TATTAATGTT	5280
45	AAATTGCTCA GTTAGTACCG ACATAACTGG TGTAGGAAAC GTTAATAATA TTAAATCTGG	5340
	CTTTCTTTT ATCAATAAAT TAATTAACCT ATTAAGTCCA TAGTATTTGT AAAAACATTT	5400
	GTCTAGTTTA TCTGGGCGGC TGTAATAAAA CCCTTTGTAC ATATTCTAA AATATTTAAA	5460
50	GCTATTGATA TACCATTTTT TACAAATAGA AGTCAAAATT GGATGAGCTT CCATAAATAA	5520
	ATCGTGCTCA ATGACGCTTA AATGGTCTAG ATTCATATCA TTAAGTTGAT TAACGATACT	5580

55

	TTGAGTAACC ATTAATAGCC ACCCTCCGTT AGTTTGAAAA TTTTATTTAA GTGTAACCTA	5700
	TTTTACGGCA TTATAAAGA AATAAAGACG CAAAGTCGTT ACATTTATAG CAATTTTAAT	5760
5	CTATAGATGA ATTGATACAA AATAAACGT TATTTTATAA AGCAATTAT TGTCTATGT	5820
	TTTATTTGTA TATTTAAAT TATCCAGTAT ACAATTATAG CATATTTTGT GAAACAATTA	5880
	TGATATTATA CCATGTTACA AGATGGTTTT AATAATTTAA GATGAGCCAT AATTGTAAAA	5940
10	CTAATTCATA ATACCGTATG TTTTATTTTT AATAGTAGAA ATTAGAAAAT GCTGATTAGT	6000
	AGGATATAAC AGTGAAATTA TAAATTTATT AACATCAACA AAACGTGTAT AATAACATA	6060
15	TTGTAGAAAA AGGAGCGGTT CAGTTTGGAT GCAAGTACGT TGTTTAAGAA AGTAAAAGTA	6120
	AAGCGTGTAT TGGGTTCTTT AGAACACAA ATAGATGATA TCACTACTGA TTCACGTACA	6180
	GCGAGAGAAG GTAGCATTTT TGTGCTTCA GTTGATATA CTGTAGACAG TCATAAGTTC	6240
20	TGTCAAATG TAGCTGATCA AGGGTGTAAG TTGGTAGTGG TCAATAAAGA ACAATCATT	6300
	CCAGCTAACG TAACACAAGT GGTGTGCCG GACACATTAA GAGTAGCTAG TATTCTAGCA	6360
	CACACATTAT ATGATTATCC GAGTCATCAG TTAGTGACAT TTGGTGTAaC GGGTACAAAT	6420
25	GGTAAACTT CTATTGCGAC GATGATTCAT TTAATTCAAA GAAAGTTACA AAAAAATAGT	6480
	GCAATTTAG GAACTAATGG TTTCCAAATT AATGAAACAA AGACAAAAGG TGCAAATACG	6540
	ACACCAGAAA CAGTTTCTTT AACTAAGAAA ATTAAAGAAG CAGTTGATGC AGGCGCTGAA	6600
30	TCTATGACAT TAGAAGTATC AAGCCATGGC TTAGTATTAG GACGACTGCG AGGCGTTGAA	6660
	TTTGACGTTG CAATATTTTC AAATTTAACA CAAGACCATT TAGATTTTCA TGGCACAATG	6720
	GAAGCATACG GACACGCGAA GTCTTTATTG TTTAGTCAAT TAGGTGAAGA TTTGTGAAA	6780
35	GAAAAGTATG TCGTGTTAAA CAATGACGAT TCATTTTCTG AGTATTTAAG AACAGTGACG	6840
	CCTTATGAAG TATTTAGTTA TGAATTGAT GAGGAAGCCC AATTTATGGC TAAAAATATT	6900
	CAAGAATCTT TACAAGGTGT CAGCTTTGAT TTTGTAACGC CTTTGGGAAC TTACCCAGTA	6960
40	AAATCGCCTT ATGTTGGTAA GTTTAATATT TCTAATATTA TGGCGGCAAT GATTGCGGTG	7020
	TGGAGTAAAG GTACATCTTT AGAAACGATT ATTAAAGCTG TTGAAAATT AGAACCTGTT	7080
45	GAAGGGCGAT TAGAAGTTTT AGATCCTTCG TTACCTATTG ATTTAATTAT CGATTATGCA	7140
	CATACAGCTG ATGGTATGAA CAAATTAATC GATGCAGTAC AGCCTTTTGT AAAGCAAAAG	7200
	TTGATATTTT TAGTTGGTAT GGCAGGCGAA CGTGATTTAA CTAAAACGCC TGAAATGGGG	7260
50	CGAGTTGCCT GTCGTGCAGA TTATGTCATT TTCACACCGG ATAATCCGGC AAATGATGAC	7320
	CCGAAAATGT TAACGGCAGA ATTAGCCAAA GGTGCAACAC ATCAAACTA TATTGAATTT	7380

55

	GTTTTAGCAT CAAAAGGAAG AGAACCATAT CAAATCATGC CAGGGCATAT TAAGGTGCCA	7500
	CATCGAGATG ATTTAATTGG CCTTGAAGCA GCTTACAAAA AGTTCGGTGG TGGCCCTGTT	7560
5	GATTAATAAA AGATTTATTG ATGAAGGTAA AACTATTGAT GTTTATTTAT TCGAAGCATT	7620
	AAATAACCAG ATAATCATTG CTATACCAGA TTGGTTTTGG TCATATCAGA TGGCAATGAC	7680
	ATTAGATGAA GAAACTTGTT TTGAAGCAAT ACTCATGCAA TTGTTTGTTT TTAAAGAAGA	7740
10	GGAAGAGGCA GAATCGATTG CATCACAAC AACAGATTGG ATAGAAACAT ATAAAAAGGA	7800
	GAAAGACTAA TGAACCTAAA GCAAGAAGTT GAGTCTAGAA AGACTTTTGC GATTATTTCA	7860
	CATCCCGATG CAGGGAAAAC AACGTTAACT GAAAACTAT TGTACTTCAG TGGTGCTATT	7920
15	CGTGAAGCGG GTACAGTTAA AGGGAAGAAG ACTGGTAAAT TTGCGACAAG TGAATGGATG	7980
	AAAGTTGAAC AAGAGCGTGG TATTTCTGTA ACTAGTTCAG TAATGCAATT TGATTACGAT	8040
20	GATTATAAAA TCAATATCTT AGATACACCA GGACATGAAG ACTTTTCAGA AGATACGTAT	8100
	AGAACATTAA TGGCAGTTGA CAGTGCTGTC ATGGTCATAG ACTGTGCAAA AGGTATTGAA	8160
	CCACAAACAT TGAAGTTATT TAAAGTTTGT AAAATGCGTG GTATTCCAAT CTTTACATTC	8220
25	ATTAATAAAT TAGACCGAGT AGGTAAAGAA CCATTTGAAT TATTAGATGA AATCGAAGAG	8280
	ACATTAAATA TTGAAACATA CCCTATGAAT TGGCCAATTG GTATGGGACA AAGTTTCTTT	8340
	GGCATCATTG ATAGAAAGTC TAAACAATT GAACCATTGA GAGATGAAGA AAATATATTA	8400
30	CATTTGAATG ATGATTTTGA GTTGAAGAA GATCATGCAA TTACAAATGA TAGTGATTTT	8460
	GAACAAGCGA TTGAAGAATT AATGTTGGTT GAAGAAGCGG GTGAAGCCTT TGATAATGAC	8520
	GCGCTGTTGA GTGGAGACTT AACACCTGTA TTTTCGGTT CAGCTTTAGC TAACTTTGGT	8580
35	GTACAAAATT TCTTAAATGC ATATGTTGAT TTGCGCCAA TGCCAAATGC GAGACAAACA	8640
	AAAGAGACG TTGAAGTAAG CCCGTTTGAT GATTCATTTT CAGGATTTAT CTTTAAAT	8700
	CAAGCCAACA TGGACCCTAA ACACCGTGAT AGAATTGCCT TTATGCGTGT CGTTAGTGGT	8760
40	GCATTTGAAC GTGGTATGGA TGTACTTTG CAACGTACTA ATAAAAAGCA AAAGATCACA	8820
	CGTTCAACGT CATTTATGGC AGACGATAAA GAACTGTGA ATCATGCTGT AGCAGGCGAT	8880
45	ATCATTGGAC TATATGATAC TGGTAATTAT CAAATTGGAG ATACTTTAGT TGGTGAAAA	8940
	CAAACCTACA GTTTCCAAGA TTTACCACAA TTTACGCCAG AAATTTTAT GAAAGTTTCT	9000
	GCTAAAAACG TCATGAAACA GAAGCATTTT CATAAAGSTA TTGAACAATT AGTACAAGAA	9060
50	GGTGCATTTC AATACTATAA AACATTACAC ACAAACCAA TTATTTTAGG TGCTGTTGGT	9120
	CAGTTACAAT TTGAAGTTTT CGAACATAGA ATGAAAAACG AATATAATGT TGATGTTGTT	9180

55



	AAGATGAACA CATCAAGATC GATTTTAGTG AAAGATAGAT ATGACGATTT AGTATTCTTA	9300
	TTTGAAAATG AATTTGCAAC AAGATGGTTT GAAGAGAAAT TCCCTGAAAT TAAATTGTAT	9360
5	AGTTTACTTT AACAGCTCAA TTGTATAATC GAATTTGTTA CATTAAAAAT AATTGTTTCG	9420
	TTGAAGAAAA ATAAATTGTA TATTTTAAAA GAAAAAGGTA TACTATGATG TATCAAATGA	9480
	ATAACCTATG GCATTTTGTC AGAGGGGAGT AACTTAAGAA TCATGACCGT ATAAATGaTT	9540
10	CGACACTTTA TCGTCATTAC GAAGATATCT TCCGGTAAAG TGGGCAATTT AAATTGCTTA	9600
	GTGAGACCTT TGCTATTTAT TTAGCATAGG TCTTTTTGTT TGTACTTAAC TTATTTATTT	9660
15	AAAGGAGTTG TACATGTTAA TGGATCCAAG TTTGATCTTA CCTTATTTAT GGGTACTTGT	9720
	CGTTTTAGTA TTTTLAGAAG GCTTATTAGC AGCAGATAAC GCGATTGTTA TGGCTGTAAT	9780
	GGTTAAGCAC TTACCACCCG AACAAACGTAA AAAAGCTTTG TTTTACGGTT TGTTAGGTGC	9840
20	ATTTGTATTT AGATTTTTAG CATTATTCTT AATTAGTATT ATCGCGAACT TTTGGTTTAT	9900
	TCAAGCTGCA GGAGCGGTTT ACTTAATTTA TATGTCAATC AAAAATCTGT GGCAGTTCTT	9960
	TAAACACCCA GAAATTGAAA GTCCTGAAGC TGGAGATGAT CATCATTATG ATGAATCTGG	10020
25	TGAAGAGATT AAAGCAAGTA ACAAATCATT CTGGGGAAC TGTGTGAAAA TAGAATTTGC	10080
	AGATATCGCA TTTGCCATTG ATTCTATGCT TGCTGCTTTA gCTATTGCTG TAACACTTCC	10140
	TAAAGTTGGT ATTCACTTTG GTGGTATGGA CTTAGGTCAG TTCGTAGTCA TGTTCTTAGG	10200
30	TGGAATGATT GGTGTTATTC TAATGCGTTA TGCAGCAACA TGGTTTGTAG AGCTATTAAA	10260
	CAAATATCCA GGACTTGAAG GTGCAGCCTt CGCGATCGTT GGTGGGTAG GTGTAAAT	10320
35	AGTTGTCATG GTATTAGCGC ACCCAGACAT CGCTGTATTG CCTGAGCACT TCCCACATGG	10380
	CGTATTATGG CAATCTATTT TCTGGACAGT ACTAATTGGA TTAGTAATTA TCGGTTGGTT	10440
	AGGTTCAGTT GTTAAAAATA AAAAATCGCA TAAATAATTG ATGTGAAGCG GACAATCTTA	10500
40	ATTTAGTTTA AGGTTGTCCT TTTTCATTTA ATTGAGTGAT TTATGAAAAA TGGATTTTGA	10560
	AGAATGTGAA TCAAAGATG CGATATAGTA TTAAGAAAAT GTGCCTTTTA TATTTAGCAT	10620
	TTTTTCAATA GAAATTATAT AGATTTTAAA GCAAATTAGG TGTTAATGTG TCATAATGAT	10680
45	AAGTGATTTT ATTGAATGGA GTGGACATTA GTGGATATTG GTAAAAACA TGTAATTCCT	10740
	AAAAGTCAGT nACCsaCGTA AGCGTCGTGA ATTCTTCCAC AACGAAGACA GAGAAGAAAA	10800
50	TTTAAATCAA CATCAAGATA AACAAAATAT AGATAATACA ACATCAAAAA AAGCAGATAA	10860
	GCAAATACAT AAAGATTCAA TTGATAAGCA CGAACGTTTT AAAAATAGTT TATCATCGCA	10920
	TTTAGAACAG AGAAACCGTG ATGTTAATGA GAATAAGCT GAAGAAAGTA AAAGTAATCA	10980
55		

	AAATTCATTA GATTCAGTGG ACCAAGATAC AGAGAAATCA AAATATTATG AGCAAAATTC	11100
	TGAAGCGACT TTATCAACTA AATCAACCGA TAAAGTAGAA TCAACTGAAA TGAGAAAGCT	11160
5	AAGTTCAGAT AAAAACAAG TTGGTCATGA AGAGCAACAT GTACTTTCTA AACCTTCAGA	11220
	ACATGATAAA GAGACTAGAA TTGATTCTGA GTCTTCAAGA ACTGATTCAG ACAGCTCGAT	11280
	GCAGACAGAG AAAATAAAAA AAGACAGTTC AGATGGAAAT AAAAGTAGTA ATCTGAAATC	11340
10	TGAAGTAATA TCAGACAAAT CAAATACAGT ACCAAAATTG TCGGAATCTG ATGATGAAGT	11400
	AAATAATCAG AAGCCATTAA CTTTACCGGA AGAACAGAAA TTGAAAAGAC AGCAAAGTCA	11460
15	AAATGAGCAA ACAAAAACCT ATACATATGG TGATAGCGAA CAAAATGACA AGTCTAATCA	11520
	TGAAAATGAT TTAAGTCATC ATATACCATC GATAAGTGAT GATAAAGATA ACGTCATGAG	11580
	AGAAAATCAT ATTGTTGACG ATAATCCTGA TAATGATATC AATACACCAT CATTATCAAA	11640
20	AACAGATGAC GATCGAAAAC TTGATGAAAA AATTCATGTT GAAGATAAAC ATAAACAAAA	11700
	TGCAGACTCG TCTGAAACGG TGGGATATCA AAGTCAGTCA ACTGCATCTC ATCGTAGCAC	11760
	TGAAAAAGA AATATTTCTA TTAATGACCA TGATAAATTA AACGGTCAAA AAACAAATAC	11820
25	AAAGACATCG GCAAATAATA ATCAAAAAAA GGCTACATCA AAATTGAACA AAGGGCGCGC	11880
	TACGAATAAT AATTATAGTG ACATTTTGAA AAAGTTTGG ATGATGTATT GGCCTAAATT	11940
	AGTTATTCTA ATGGGTATTA TTATTCTAAT TGTTATTTTG AATGCCATTT TTAATAATGT	12000
30	GAACAAAAT GATCGCATGA ATGATAATAA TGATGCAGAT GCTCAAAAAT ATACGACAAC	12060
	GATGAAAAT GCCAATAACA CAGTTAAATC GGTCGTTACA GTTGAAAATG AAACATCAAA	12120
35	AGATTCmTCA TTACCTAAAG ATAAAGCATC TCaAGACGAA GTGGGATCAG GTGTTGTATA	12180
	TAAAAAATCT GGAGATACGT TATATATTGT TACGAATGCA CACGTTGTCG GTGATAAAGA	12240
	AAATCaAAAA ATAACCTTCT CGAATAATAA AAGTGTTGTT GGGAAAGTGC TTGGTAAAGA	12300
40	TAAATGGTCA GATTTAGCTG TTGTTAAAGC AACTTCTTCA GACAGTTCAG TGAAAGAGAT	12360
	AGCTATTGGA GATTCAAATA ATTTAGTGTT AGGAGAGCCA ATATTAGTCG TAGGTAATCC	12420
	ACTTGGTGTA GACTTTAAAG GCACTGTGAC AGAAGGTATT ATTTCAGGTC TGAACAGAAA	12480
45	TGTTCTTATT GATTTGATA AAGATAATAA ATATGATATG TTGATGAAAG CTTTCCAAAT	12540
	TGATGCATCA GTAAATCCAG GTAACCGGG TGGTGCTGTC GTCAATAGAG AAGGAAAATT	12600
50	AATAGGTGTA GTTGCAGCTA AAATTAGTAT GCCAACGTT GAAAnTATGT CATTTGCA	12658

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6048 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

	TGAAATnGAA TAGTACTATT GCAAGTGTAAGAGGTTAAT TTTTGCCnCA CGCGGGACTT	60
10	AAAAAGGCAA CCACTGGTTG TGACATATCC TTATTTACAT TTATAAATAT AAGGAGGAGG	120
	TAGTAGTGAA AGACTTATTG CAAGCACAGC AAAAGCTTAT ACCGGATCTC ATAGATAAAA	180
	TGTATAAACG TTTTCTATT CTTACTACTA TCTCAAAAAA TCAGCCTGTC GGACGTCGAA	240
15	GTTTAAGCGA ACATATGGAT ATGACTGAAC GTGTACTGCG TTCTGAAACA GATATGCTTA	300
	AGAAACAAGA TTTGATAAAA GTTAAGCCTA CCGGAATGGA AATTACAGCT GAAGGTGAGC	360
	AACTGATTTC GCAATTGAAA GGTACTTTG ATATCTATGC AGATGATAAT CGTCTGTCAG	420
20	AAGGTATTAA GAATAAATTT CAAATTAAGG AAGTTCATGT TGTTCTGGT GATGCTGATA	480
	ATAGTCAATC TGTTAAACA GAATTAGGTA GACAAGCAGG TCAATTACTT GAAGGCATAT	540
	TACAAGAAGA CGGATAGTT GCTGTAACG GCGATCCAC GATGGCATGT GTTAGTGAAG	600
25	CAATTCATTT ATTACCATAT AATGTATTCT TCGTACCAGC CAGAGGTGGA CTAGGCGAAA	660
	ATGTTGTCTT TCAGGCAAAC ACAATTGCAG CCAGTATGGC aCAACAAGCT GGCGGTTATT	720
30	ATACGACGAT GTATGTACCT GATAATGTCA GTGAAaCAAC ATATAATACA TTGTTGTTAG	780
	AGCCATCAGT CATAAACACT TTAGACAAAA TTAACAAGC AAACGTTATA TTACACGGCA	840
	TTGGTGATGC GCTGAAGATG GCGCATCGAC GTCAATCACC TGAAAAGGTC ATTGAACAAC	900
35	TTCAACATCA TCAAGCTGTC GGAGAGGCAT TTGGTTATTA TTTTGATACA CAAGGTCAAA	960
	TTGTCCATAA GGTTAAACA ATTGGACTTC AATTAGAAGA CCTTGAATCA AAAGACTTTA	1020
	TTTTTGAGT TGCAGGAGGC AAATCGAAAG GTGAAGCAAT TAAAGCATACTTGACGATTG	1080
40	CACCCAAGAA TACAGTGTTA ATCACTGATG AAGCCGCAGC AAAGATAATA CTTGAATAAG	1140
	AGATAAAAAG TTTAATACTT TTTAAATATC ATTTTAAAGG AGGCCATTAT AATGGCAGTA	1200
	AAAGTAGCAA TTAATGGTTT TGGTAGAATT GGTGTTTATG CATTGAGAAG AATTCAAGAA	1260
45	GTAGAAGGTC TTGAAGTTGT AGCAGTAAAC GACTTAACAG ATGACGACAT GTTAGCGCAT	1320
	TTATTAAAAT ATGACACTAT GCAAGGTCGT TTCACAGGTG AAGTAGAGGT AGTTGATGGT	1380
50	GGTTTCCGCG TAAATGGTAA AGAAGTTAAA TCATTCAGTG AACCAGATGC AAGCAAATTA	1440
	CCTTGGAAG ACTTAAATAT CGATGTAGTA TTAGAATGTA CTGGTTTCTA CACTGATAAA	1500
	GATAAAGCAC AAGCTCATAT TGAAGCAGGC GCTAAAAAAG TATTAATCTC AGCACCAGCT	1560

55

	ACAGTTGTTT CAGGTGCTTC ATGTACTACA AACTCATTAG CACCAGTTGC TAAAGTTTTA	1680
	AACGATGACT TTGGTTTAGT TGAAGGTTTA ATGACTACAA TTCACGCTTA CACAGGTGAT	1740
5	CAAAATACAC AAGACGCACC TCACAGAAAA GGTGACAAAC GTCGTGCTCG TGCAGCGGCA	1800
	GAAAACATCA TCCCTAACTC AACAGGTGCT GCTAAAGCTA TCGGTAAAGT TATTCCTGAA	1860
10	ATCGATGGTA AATTAGATGG TGGTGCACAA CGTGTTCTTG TAGCTACAGG TTCATTAACT	1920
	GAATTAACAG TAGTATTAGA AAAACAAGAC GTAACAGTTG AACAAGTTAA CGAAGCTATG	1980
	AAAAATGCTT CAAACGAATC ATTCCGGTtAC ACTGAAGACG AAATCGTTTC TTCAGACGTT	2040
15	GTAGGTATGA CTTACGGTTC ATTATTCGAC GCTACACAAA CTCGTGTAAT GTCAGTTGGC	2100
	GACCGTCAAT TAGTTAAAGT TGCAGCTTGG TATGATAACG AAATGTCATA TACTGCACAA	2160
	TTAGTTCGTA CATTAGCATA CTTAGCTGAA CTTTCTAAAT AATTTTAGTA TAGTTTTTAT	2220
20	TCAAATACGC TAGTGCTCAG AACTATTTAG CATTAAATTAA AGCTTATGAG TAAGCGGGGA	2280
	GCACAAACGC TTCTCCGCTT ATTTTTATAT AAAATTTCTT AATTACAAGG AGGAAACACC	2340
	ATGGCTAAAA AAATTGTTTC TGATTTAGAT CTTAAAGGTA AAACAGTCCT AGTACGTGCT	2400
25	GATTTTAACG TACCTTTAAA AGACGGTGAA ATTACTAATG ACAACCGTAT CGTTCAGCT	2460
	TTACCTACAA TTCAATACAT CATCGAACAA GGTGGTAAAA TCGTACTATT TTCACATTTA	2520
30	GGTAAAGTGA AAGAAGAAAG TGATAAAGCA AAATTAACCT TACGTCCAGT TGCTGAAGAC	2580
	TTATCTAAGA AATTAGATAA AGAAGTTGTT TTCGTACCAG AAACACGCGG CGAAAACTT	2640
	GAAGCTGCTA TTAAAGACCT TAAAGAAGGC GACGTATTAT TAGTTGAAAA TACACGTTAT	2700
35	GAAGATTTAG ACGGTAAAAA AGAATCTAAA AATGATCCAG AATTAGGTAA ATACTGGGCA	2760
	TCTTTAGGTG ATGTGTTTGT AAATGATGCT TTTGGTACTG CGCATCGTGA GCATGCATCT	2820
	AATGTTGGTA TTTCTACACA TTTAGAACT GCAGCTGGAT TCTTAATGGA TAAAGAAAT	2880
40	AAGTTTATTG GCGGCGTAGT TAACGATCCA CATAAACCAG TTGTTGCTAT TTTAGGTGGA	2940
	GCAAAAGTAT CTGACAAAAT TAATGTCATC AAAAAGTTAG TTAACATAGC TGATAAAATT	3000
	ATCATCGGCG GAGGTATGGC TTATACTTTC TTAAAAGCGC AAGGTAAAGA AATTGGTATT	3060
45	TCATTATTAG AAGAAGATAA AATCGACTTC GCAAAAGATT TATTAGAAAA ACATGGTGAT	3120
	AAAATTGTAT TACCAGTAGA CACTAAAGTT GCTAAAGAAT TTTCTAATGA TGCCAAAATC	3180
50	ACTGTAGTAC CATCTGATTC AATTCCAGCA GACCAAGAAG GTATGGATAT TGGACCAAAC	3240
	ACTGTAAAAT TATTTGCAGA TGAATTAGAA GGTGCGCACA CTGTTGTATG GAATGGACCT	3300
	ATGGGTGTAT TCGAGTTCAG TAACTTTGCA CAAGGTACAA TTGGTGTATG TAAAGCAATT	3360
55		

	TCTTTAGGTT	TTGAAAATGA	CTTCACTCAT	ATTTCAACTG	GTGGCGGCGC	GTCATTAGAG	3480
	TACCTAGAAG	GTAAAGAATT	GCCTGGTATC	AAAGCAATCA	ATAATAAATA	ATAAAGTGAT	3540
5	AGTTTAAAGT	GATGTGGCAT	GTTTGTTTAA	CATTGTTACG	GGAAAACAGT	CACAAGATGA	3600
	CATCGTGTTC	CATCACTTTT	CAAAAATATT	TACAAAACAA	GGAGTGCTTT	TAATGAGAAC	3660
	ACCAATTATA	GCTGGTAACT	GGAAAATGAA	CAAAACAGTA	CAAGAAGCAA	AAGatTCGTC	3720
10	AATACATTAC	CAACACTACC	AGATTCAAAA	GAAGTAGAAT	CAGTAATTTG	TGCACCAGCA	3780
	ATTCAATTAG	ATGCATTAAAC	TACTGCAGTT	AAAGAAGGAA	AAGCACAAGG	TTTAGAAATC	3840
15	GGTGCTCAAA	ATACGTATTT	CGAAGATAAT	GGTGC GTTCA	CAGGTGAAAC	GTCTCCAGTT	3900
	GCATTAGCAG	ATTTAGGCGT	TAAATACGTT	GTTATCGGTC	ATTCTGAACG	TCGTGAATTA	3960
	TTCCACGAAA	CAGATGAAGA	AATTAACAAA	AAAGCGCACG	CTATTTTCAA	ACATGGAATG	4020
20	ACTCCAATTA	TATGTGTTGG	TGAAACAGAC	GAAGAGCGTG	AAAGTGGTAA	AGCTAACGAT	4080
	GTTGTAGGTG	AGCAAGTTAA	GAAAGCTGTT	GCAGGTTTAT	CTGAAGATCA	ACTTAAATCA	4140
	GTTGTAATTG	CTTATGAACC	AATCTGGGCA	ATCGGAAGTG	GTAAATCATC	AACATCTGAA	4200
25	GATGCAAATG	AAATGTGTGC	ATTTGTACGT	CAAACTATTG	CTGACTTATC	AAGCAAAGAA	4260
	GTATCAGAAG	CAACTCGTAT	TCAATATGGT	GGTAGTGTTA	AACCTAACAA	CATTAAAGAA	4320
	TACATGGCAC	AAACTGATAT	TGATGGGGCA	TTAGTAGGTG	GCGCATCACT	TAAAGTTGAA	4380
30	GATTTTCGTAC	AATTGTTAGA	AGGTGCAAAA	TAATCATGGC	TAAGAAACCa	ACTGCGTTAA	4440
	TTATTTTAGA	TGGTTTTGCG	AACCGCGAAA	GCGAACATGG	TAATGCGGTA	AAATTAGCAA	4500
35	ACAAGCCTAA	TTTTGATCGT	TATTACAACA	AATATCCAAC	GAATCAAATC	GAAGCGAGTG	4560
	GCTTAGATGT	TGGACTACCT	GAAGgACAAA	TGGGTAACTC	AGAAGTTGGT	CATATGAATA	4620
	TCGGTGCAGG	ACGTATCGTT	TATCAAAGTT	TAACTCGAAT	CAATAAATCA	ATTGAAGACG	4680
40	GTGATTTCTT	TGAAAATGAT	GTTTTAAATA	ATGCAATTGC	ACACGTGAAT	TCACATGATT	4740
	CAGCGTTACA	CATCTTTGGT	TTATTGTCTG	ACGGTGGTGT	ACACAGTCAT	TACAAACATT	4800
	TATTTGCTTT	GTTAGAACTT	GCTAAAAAAC	AAGGTGTTGA	AAAAGTTTAC	GTACACGCAT	4860
45	TTTTAGATGG	CCGTGACGTA	GATCAAAAAT	CCGCTTTGAA	ATACATCGAA	GAGACTGAAG	4920
	CTAAATTCAA	TGAATTAGGC	ATTGGTCAAT	TTGCATCTGT	GTCTGGTCGT	TATTATGCAA	4980
	TGGATCGTGA	CAAACGTTGG	GAACGTGAAG	AAAAAGCTTA	CAATGCTATT	CGTAATTTTG	5040
50	ATGCCCCAAC	TTATGCAACT	GCCAAAGAAG	GTGTAGAAGC	AAGCTATAAT	GAGGGCTTAA	5100
	CTGACGAATT	CGTAGTACCA	TTCATCGTTG	AGAATCAAAA	TGACGGTGTT	AATGATGGAG	5160
55							

CGAACAGAGC ATTCGAAGGC TTAAAGTTG AACAAAGTTAA AGACTTATTC TATGCAACAT 5280  
TCACTAAGTA TAATGACAAT ATCGATGCGG CTATCGTCTT CGAAAAAGTT GATTTAAATA 5340  
5 ATACAATTGG TGAAATTGCA CAAAATAACA ATTTAACTCA ATTACGTATT GCAGAACTG 5400  
AAAAATACCC TCACGTTACT TACTTTATGA GTGGTGGACG TAACGAGGAA TTAAAGGTG 5460  
AACGCCGTCG TTAAATTGAT TCACCTAAAG TTGCAACGTA TGACTTGAAA CCAGAAATGA 5520  
10 GTGCTTATGA AGTTAAAGAT GCATTATTAG AAGAGTTAAA TAAAGGTGAC TTGGACTION 5580  
TTATTTTAAA CTTTGCTAAC CCTGATATGG TTGGACATAG TGGTATGCTT GAGCCGACAA 5640  
TCAAAGCAAT CGAAGCGGTT GATGAATGTT TAGGAGAAGT GGTGATAAG ATTTTAGACA 5700  
15 TGGACGGTTA TGCAATTATT ACTGCTGACC ATGGTAACTC TGATCAAGTA TTGACGGaTG 5760  
ATGATCAACC AATGACTACG CAWACAACGA ACCCAGTACC AGTGATTGTA ACAAAGAAG 5820  
20 GCGTTACACT TAGAGAACT GGTCGCTTAG GTGACTTAGC ACCTACATTA TTAGATTTAT 5880  
TAAATGTAGA ACAACCTGAA GATATGACAG GTGAaTCTTT AATTAAACAC TAATATTGTA 5940  
AAAGATGTTA AGTAAACGCT TAATGACACT TATTTTTTGA AAATAATAGT AATATChTTT 6000  
25 TGTAAATGA AAGAATAAAG CTATAATAAT TATAGAATAA CTATTTAn 6048

## (2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:  
30 (A) LENGTH: 5602 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

AAAGaAGTGC AAGATATCAT CGCATTAAAT AAGTCGTTAC AAaGTGTAAT TGTAGACaTC 60  
40 GCTTCCAATA ATGTTGATAC AATTATGCCT GGTATACTC ATTTACAGCG TGCACAGCCA 120  
ATTTCAATTG CACATCATAT TATGACTTAT TTTTGGATGT TACAACGAGA CCAACAACGA 180  
TTTGAAGATA GTTTAAAACG AATCGATATT AATCCTTTAG GTGCAGCAGC CTTAAGTGGT 240  
45 ACCACATACC CTATCGATAG ACACGAGACA ACAGCATTGT TGAACCTTGG CAGTCTCTAT 300  
GAGAATAGCC TAGATGCTGT TAGTGACAGA GACTATATTA TTGAAACATT GCATAATATT 360  
TCTTTAACGA TGGTTCACCT ATCACGCTTT GCAGAGGAAA TTATTTTCTG GTCCACAGAC 420  
50 GAAGCTAAAT TCATTACATT ATCAGATGCA TTTTCAACTG GCTCATCTAT TATGCCACAA 480  
AAGAAAAATC CTGATATGGC AGAATTAATT AGAGGTAAAG TTGGTCGAAC GACTGGTCAT 540

55

## EP 0 786 519 A2

	GAAGATAAAG AAGGTTTATT CGATGCTGTC CATACAATTA AAGGTTCTTT ACGTATTTTC	660
	GAAGGTATGA TTCAAACGAT GACAATTAAT AAAGAACGAC TCAATCAAAC TGTTAAAGAA	720
5	GATTTTTCAA ATGCAACGGA ACTAGCAGAT TATTTAGTAA CTAAAAATAT TCCATTTAGA	780
	ACTGCACATG AAATTGTAGG AAAAATCGTC TTAGAATGTA TACAACAAGG TCATTATTTA	840
10	TTAGATGTTT CTTTAGCAAC ATATCAACAA CATCATTCTA GTATTGATGC CGATATTTAC	900
	GATTATTTGC AGCCTGAAAA TTGTTTAAAA CGACGTCAA GTTACGGTTC AACAGGTCAA	960
	TCATCGGTCA AACAACAAC TGATGTTGCT AAACAATTAC TATCACAATA AATACGTTAA	1020
15	TCTACCTACC CACAATGTCT ATTTAAATTA CATTGTGGGT ATTTTAATGC TCTCTTCGTC	1080
	TTGTTGAACA TCACATTTTT AAGATTCCTA AAATGTTTGA TAATTCTTTT AAATTTATAT	1140
	TACAAAAATG TTATAAATG TAAAAGAAAT GTGTAAAGCG TTTTCACAAG CAGGTTTTTG	1200
20	TAGTATTTTA AAATTGTTAG ACTACAAATA AAGAGATGAA AGGATAAAGA CTATGACTAA	1260
	CTCTTCGAAA AGCTTCACTA AATTTATGGC TGCTTCTGCT GTTTTACTA TGGGATTTTT	1320
	ATCAGTACCT ACTGCTGGCG CTGAACAAAC AAATCAAATT GCAAATAAAC CTCAGGCTAT	1380
25	TCAATGGCAT ACAAATTTAA CGAATGAGCG ATTCACTACT ATCGCACATC GTGGCGCAAG	1440
	TGGCTATGCA CCCGAGCATA CGTTTCAAGC ATATGATAAG AGTCATAATG AGTTAAAAGC	1500
30	ATCTTATATC GAAATTGATT TACAACGTAC CAAAGATGGC CATTTAGTTG CTATGCATGA	1560
	TGAAACTGTT AACCGTACAA CAAATGGACA CGGTAAAGTT GAGGATTATA CCCTTGATGA	1620
	ATTTAAACAG TTAGATGCAG GAAGTTGGTT TAATAAAAAA TATCCAAAAT ACGCAAGAGC	1680
35	AAGTTATAAA AATGCTAAAG TACCCACTTT AGATGAAATT TTAGAACGTT ATGGCCCGAA	1740
	TGCAAACTAT TATATTGAAA CAAAGTCACC TGATGTATAC CCAGGAATGG AAGAACAATT	1800
	ATTAGCTTCA TTGAAAAAGC ATCACCTTTT AAATAACAAT AAATTAAAAA ATGGACATGT	1860
40	AATGATTCAA TCATTTTCTG ACGAAAGTTT AAAGAAAATT CATCGTCAAA ATAAGCATGT	1920
	GCCATTAGTA AAATTAGTTG ATAAAGGTGA ACTACAACAA TTTAACGACC AACGCTTAA	1980
45	AGAGATACGC TCTTATGCGA TTGGATTAGG TCCTGATTAT ACAGATTTAA CTGAACAAAA	2040
	TACCCATCAT TTAAAAGACT TAGGATTTAT AGTACATCCT TATACAGTGA ATGAAAAAGC	2100
	TGATATGTTA CGATTAAATA AATATGGCGT TGATGGTGTC TTTACAAATT TCGCTGATAA	2160
50	ATATAAGAA GTCAITTAAGT AGTAATGTTA AACTAGAAAA CATAAATACA AAAATATAGC	2220
	TATTACTATA AAAACAGCA GTAAGATATT TCCAAATTGA AATTATCCTA CTGCTGTCTT	2280
55	TTTGGGAGTG GGACAGAAAT GATATTTTCG CAAAATTTAT TTCGTCGTCC CACCCCAACT	2340

	TTGTCTGTAG AAATTGAGGA GCTAATTTCT CTGTGTCGGG GCTCCACCCC AACTTGCACA	2460
	CTATTGTAAG CTGACTTTCC GCCAGCCTCT GTGTGCGGGC CCCGCCAACT TGCACACTAT	2520
5	TGTAAGCTGA CTTTCCACCA GCCTCTGTGT TGGGGCCCCG ACTATTTTGT AAAAGAGCGT	2580
	GTTACACGGG CATTGTTTTA CAGTCAACTA CTGCTAAAAT AAAATTAACG AGCTTAGGGC	2640
10	TTTGTTTTCT GTCCCAAGCT CGTTAAATCA CATATGATAA TTAATTATGC CCAACCACGA	2700
	TATCTAGCTG CTTCTGCTGT ACGTTTAATA CCTATGATAT ATGCTGCAAG TCTCATATCT	2760
	ATTTTTCGGT TTTGAGACAA TTCGTAAATC GTATCAAATG CCGCTTCTAA TTTTTCACGT	2820
15	AGCTTTTCAT TAACTTCTTC TTCAGACCAA TAATAACCTT GATTATTTTG TACCCATTCTG	2880
	AAGTAAGAAA CCGTACACACC ACCAGCACTT GCTAATACGT CTGGAACATA TAATATACCA	2940
	CGTTCAGTTA AAATACGTGT TGCTTCTGGT GTTGTAGGTC CATTAGCAGC TTCAACAACG	3000
20	ATACTAGCTT TAATATCATG TGCATTGTCT TCTGTAATTT GGTGTGAAAT AGCCGCTGGT	3060
	ACTAAAATGT CACAATCTAA TTCAAACAAT TCTTTATTTG AGATTGTTTC TTCAAATAAA	3120
	TTTGTTACCG TACCAAAACT ATCACGACGG TCTAATAAAT AATCTATATC TAAGCCATTT	3180
25	GGATCGTGTA ATGCACCGTA AGCATCAGAG ATACCTACAA TTTTTCACC TAAATCATAT	3240
	AAGAATTTAG CTAAGAAACT TCCGGCATTG CCGAAACCTT GAATAACAAC CTGGGCACCT	3300
30	TCAATTGCA TATTACGACG TTTTGCAGCT TGTTC AATTG CAATAACTAC ACCTAGTGCA	3360
	GTTGATCTGT CGCGTCCATG AGAACCACCC AATACAATTG GTTTACCTGT GATGAAACCT	3420
	GGTGAATTAA ATTTATCTAA TGCACTATAT TCATCCATCA TCCAAGCCAT AATTTGTGAG	3480
35	TTTGTAAATA CATCTGGTGC TGGAAATATCT TTGTTCCGAC CTACGAATTG TGAAATTGCT	3540
	CTTACATATC CGCGTGATAA ACGTTCAACT TCATGAATGC TCATTTGACG TGGATCACAA	3600
	ACGATACCAC CCTTACCACC ACCGTATGGT AAGTTTACAA TGCCACATTT CAAAGTCATC	3660
40	CACATTGATA ATGCTTTTAC TTCTTCTTCA TCAACATCTG GGTGGAAACG CACGCCCCCT	3720
	TTTGTGGTGC CAACAGCATC ATTATGTTGC GCACGGTAAC CTGTGAATGT TTTTACTGTG	3780
	CCATCATCCA TTCGTACAGG GATACGCACT TGTAACATTC TTAAAGGTTT TTTAATTAAA	3840
45	TCGTACATTC CTTCGTCAAA TCCCAATTTA TGCAATGCTT CTTTAATAAT TCCTTGAGTA	3900
	GAAGTTACTA AATTATGTG CTCAGTCATG ATCCTTTTCG CCTCTTCTTT ACCTAATGAT	3960
50	TTGCTTTTCA AACATATTGT AACATAACGT ATTCTTTT AAAGCCCTTA CAAACTGATT	4020
	GTTACAACCT TTTGACATTA TTGAAATACA TGTCTTATTT TTTCAAGTGC AAGGTCCAAT	4080
55	TCTTCTTTAG TAATAATTAA TGGTGGTGCA AAACGAATGA CAGTATCATG CGTTTCTTTA	4140



ACACCTATAA ACAAACCACG TCCACGGACT TCTTTAATTG ATGGATGATC AATTTGCTTT 4260  
 AATTGTTCTT TAAAATAATC TCCTAATTCT AAAGAGCGGC CTGGTAAATC CTCATCAACG 4320  
 5 ATAACATCTA ATGCAGCAAT TGATGCAGCA CAAGCAAGTG GATTACCACC AAATGTTGAA 4380  
 CCATGTGAGC CAGGTGTAAA GACATCTAAT ACTTCTTTAT CTGCTAATAC AACAGAAATT 4440  
 GGGAAAGACTC CACCACCTAG TGCTTTACCT AAAATATAGA CATCAGGTTT TACATTATCC 4500  
 10 CAATCCGTAG CAAATAATTT ACCCGAACGA CCTAATCCTG CTTGGATTTC GTCAGCAATA 4560  
 AATAAGACAT TATGTTTCATC ACATAATTCT CTAATTGCTT TCAAATATCC TTCTGGCGGT 4620  
 ATATTTATAC CCGCTTCACC TTGAATTGGT TCTACTAAAA CTGCTGCAGT ATTTTCATTA 4680  
 15 ATTGCAGCTT TCAATGCATC TACATCTCCA AAATCAACTT TTCTAAATCC ATCTAATAAC 4740  
 GGACCATAAC CACGTTGGTA TTCTGCTTCT GAAGATAATG AAAGTGGCGC CATTGTTCTGA 4800  
 CCATGGGAAGT TACCATTAAA TGCAATGATT TCTGCTTTAT TTGGCTCAAT TCCTTTAACA 4860  
 20 TCGTATGCCC AGCGTCGTGC TGCTTTCAAA GCTGTTTCTA CTGCTTCAGC ACCTGTATTC 4920  
 ATTGGTAAAG CTTTATCTTT ACCTGCCAGT TTACAAATTT TTTCGTACCA TTCACCTAAG 4980  
 25 TTATCACTAT GAAAAGCAGG TGAACTAAA GTCACTTTAT CAGCTTGATC TTTTAATGCT 5040  
 TGAATAATTT TCGGATGTCT ATGACCTTGG TTAACAGCGG AATATGCAGA TAACATATCC 5100  
 ATATATTTAT TGCCTTCAGG ATCTTTAACC CATACCCCTT CAGCTTcTGa AATGaCAATT 5160  
 30 GGcAATGGTA AATAATTATG TGCTCCGTAA TGATTTGTTA ACTCAATAAT TTTTTCAGAT 5220  
 TTAGTCATCA TATCTCCCCT TTTCATCATT TATAACTATT ATACATGAAA CATTATCCAA 5280  
 ATAATTACAT TAGTTTTCAA AGCAGATACT TTTCCACCAA AAAAGATGAA ATAATCACTA 5340  
 35 AGTTTCATTA AATTTGTCTA TTTTGAAAAC CCTTACATTT ATAATGACAT AATTACTTAA 5400  
 ATGaTTACAA GCAAAAGAAT TGATAATTTT ACACTTAATC AAAAGTATAT TTTACTAAGA 5460  
 40 ATATTTTAT TTATAAATAT TGAAAACCAC TAACAAATTG CATACACAAT ATCATTAGTG 5520  
 GTAACAGTTA AACACTTATT TATCTTTACG GGGTAATGGG TTAAAACCCT TnCATTAAAA 5580  
 TTGGATGnCC ATAAAATTAG GG 5602

45 (2) INFORMATION FOR SEQ ID NO: 130:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5924 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 50 (D) TOPOLOGY: linear

55

	TAACCCCAATT TTACCTGGAA AAATCgTTTG CGATGCaATm GCaTTtGaAT ATAAATACAT	60
	TTTACGTATa GAATTATAAA AgGTTTCATT CaAATCTTAG GGTCAAAAAT GTTATAATAT	120
5	TTTTATGTCA AATTTAAAAC AGTAACACTT ATTTACAAGG TTGCAATATT TTGAAGTAAT	180
	AAAGGAAGTG TCGCGTATTT TAACTTTTTTC AGAGCAAAAT GCACTCGCGA AAATAGATGA	240
	TTTAATGAAT ACTTATTGCA ATCAATGTCC AATCAAAACT CGTCTGCGTA AATTAGAGGG	300
10	GAAACGAAG GCGCATCATT TTTGTATCAA TGAGTGTTCA ATAGGGAAAG AAATAAAACA	360
	ATTAGGAAAT GAACTTCAAT AGGAGGAAGT CAAATGAAAA TTATATCTAT ATCAGAAACA	420
15	CCGAACCACA ACACAATGAA GATTACACTT AGTGAAAGCA GAGAAGGTAT GACATCAGAT	480
	ACGTATACTA AAGTTGATGA TTCACAGCCA GCATTTATTA ATGACATCTT AAAGGTTGAA	540
	GGCGTTAAAT CAATTTTCCA TGTTATGGAC TTTATTTTCA TAGATAAAGA AAATGACGCA	600
20	AATTGGGAAA CAGTATTGCC AAAAGTAGAG GCTGTATTCG AATAAATTTT TCATCAACTA	660
	GTATTCGGGG GGAATAAAGT ATATGGAAAT TTTACGTATA GAGCCAACAC CAAGTCCAAA	720
	TACAATGAAA GTTGTTTTGT CATATACAAG AGAAGACAAG TTATCTAATA CTTATAAAAA	780
25	AGTAGAAGAA ACACAACCAA GATTTATAAA TCAGTTGTTA TCTATAGATG GTATCACTTC	840
	CATTTTTTCAT GTCATGAACT TCTTAGCTGT TGATAAGGCA CCAAAGCTG ATTGGGAAGT	900
	CATATTACCT GATATTAAAG CTGCTTTTTTC TGATGCGAAT AAGGTTTTAG AATCTGTAAA	960
30	TGAACCTCAA ATTGACAATC ATTTTGGTGA AATTAAAGCT GAATTATTAA CTTTTAAGGG	1020
	TATACCGTAT CAAATTAAGC TAACTTCTGC TGACCAAGAA TTAAGAGAAC AATTACCACA	1080
35	AACATATGTT GACCATATGA CTCAAGCGCA AACAGCACAT GACAATATTG TTTTATGCG	1140
	TAAATGGCTA GATTTAGGAA ATCGCTATGG AAATATTCAA GAAGTAATGG ATGGTGTCTT	1200
	AGAAGAAGTG CTAGCTACCT ATCCAGAATC ACAGTTACCC GTATTGGTAA AACATGCTTT	1260
40	AGAAGAAAAT CACGCAACTA ATAATTATCA TTTCTATCGA CATGTCTCTT TGGATGAATA	1320
	TCATGCAACT GATAATTGGA AGACTCGATT ACGAATGTTA AACCATTTTC CAAAGCCGAC	1380
	TTTTGAAGAT ATACCGCTGC TTGATTTAGC TTTATCTGAT GAAAAAGTAC CGGTTAGACG	1440
45	TCAAGCGATT GTATTATTAG GTATGATTGA AAGTAAAGAA ATTTTACCGT ATTTATATAA	1500
	GGGGCTTCGT GATAAAAGTC CTGCTGTAAG AAGAACAGCA GGGGATTGCA TAAGCGATTT	1560
	AGGGTATCCA GAGGCACTAC CAGAAATGGT GCTACTATTA GATGATCCAC AGAAAATCGT	1620
50	TAGGTGGCGT GCTGCTATGT TTATCTTTGA TGAAGGTAAT GCAGAGCAGC TTCCCGCACT	1680
	AAAAGCCCAT ATTAATGACA ATGCGTTTGA AGTTAAATTA CAAATTGAAA TGGCCATATC	1740
55		

EP 0 786 519 A2

	AATTTAATTG GAGGAATTAA ATATGAATGC ATATGATGCT TATATGAAAG AAATTGCGCA	1860
	ACAAATGCGT GCGGAATTAA CTCAAAATGG TTTTACAAGT TTAGAAACGA GCGAACAGct	1920
5	ATCGGAGTAT ATGAACCAAG TAAATGCTGA TGACACTACT TTTGTAGTTA TTAACCTCTAC	1980
	ATGCGGCTGT GCAGCTGGAT TAGCAAGACC AGCTGCAGTA GCAGTTGCAA CACAAAATGA	2040
	ACATAGACCT ACAAATACAG TTACAGTTTT TGCTGGGCAA GATAAAGAAG CAACTGCTAC	2100
10	AATGCGAGAA TTCATTGAGC AAGCACCATC TAGTCCTTCG TATGCTTTAT TCAAAGGTCA	2160
	AGATTTAGTT TATTTTATGC CTAGAGAATT TATCGAAGGT AGAGATATTA ATGACATTGC	2220
15	AATGGACTTA AAGGATGCCT TTGACGAAAA TTGTAAATAG TACACATAAA TAAATATAAA	2280
	GGTTAACACA TTTTATAATA TTA AAAATGG TGTCTGTCAT TGAAAATAGA GAATATAGTT	2340
	GTATTCTATT TGT TAAATAA AGTCCGTTTT TACCaaCTAT ATTTTCTAGA AATTTAACTG	2400
20	TTTTAATAGG ACATCAAACA TAATATTCaA ATCaTGTGTT AACCTCTTTT TTAAAATTTT	2460
	TTAGCATTAA AGTTATAGAT TTGGGTAAAC AATTACCAAT TGGAAACATA TATCACGTTA	2520
	CGATGGGGTA GGTACTTAAT CAGCATTTTA TAAATAAAGT AACGGAATTC ATGATATTAA	2580
25	TATCATATTC CTAAATGAG TGATAACAAA ATGCTACATA AAGTTAAGTT ATATCAAAC	2640
	AAATATACAT ACTATAAATA ATGAAAATGA GGTGTTATCG CATATGTTGA ATTCATTGTA	2700
	TGCAGCATAT CACAGTCTTT GTGAAGAAGT TTTAGAAATA GGAAATACAC GAAATGATCG	2760
30	CACAAATACA GGTACGATTT CGAAATTTGG TCATCAACTT CGCTTTGACT TATCTAAAGG	2820
	ATTTCCACTA TTAACGACAA AGAAAGTTTC TTTTAAATTA GTAGCAACCG AATTATTATG	2880
35	GTTCAATAAA GGAGATACAA ACATCCAATA CTTATTAAAA TATAATAATA ATATATGGAA	2940
	CGAATGGGCT TTTGAAAATT ATATCAAATC AGACGAGTAT AAAGGTCCAG ATATGACAGA	3000
	TTTCGGGCAT CGTGCAATTGA GTGATCCTGA ATTTAACGAA CAATATAAAG AACAAATGAA	3060
40	ACAATTTAAG CAACGTATTC TTGAAGATGA TACATTTGCG AAGCAATTCG GGGATTTAGG	3120
	AAATGTTTAT GGTAAACAAT GGCGAGATTG GGTGATAAA GATGGTAATC ATTTTGATCA	3180
	ACTTAAACA GTAATTGAAC AAATTAAGCA TAATCCAGAT TCAAGGCGAC ACATCGTATC	3240
45	TGCATGGAAT CCAACAGAAA TTGATACAAT GGCACTTCCG CCTTGTCTATA CCATGTTCCA	3300
	GTTTTATGTC CAAGATGGTA AGTTAAGTTG CCAGTTATAC CAACGTAGCG CAGATATCTT	3360
	TTTAGGTGTG CCATTTAATA TCcGCagctA CGCTTTATTG ACACACCTTA TTGCCAAAGA	3420
50	ATGTGGACTT GAAGTGGGTG AATTGTGCA TACATTTGGA GATGCACATA TTTATTCAAA	3480
	TCATATTGAT GCGATTCAAA CACAATTAGC ACGTGAAAGC TTCAATCCTC CAACATTAAA	3540
55		

	TGAATCACAT CCAGCAATAA AAGCTCCAAT AGCAGTGTAG TCATTGCATA GTTAGCTAAC	3660
	CATATAGACA TCAAAATGAC ATCATAGTAT TTTCAAGTGC AAAAAAGTAC TTTTTGTGT	3720
5	TAAACGTTTT CATAAATTAT GCAAAATCAT TATTTCTATC ACACTTTATG ATAAAAATTG	3780
	TGTTAAATTA AAGATAACTT AGTAATAAAA AATGAAATGA TAGAAGAAGG AGGATAATTA	3840
	TGACTTTATC CATTCTAGTt GCACATGACT TGCAACGAGT AATTGGTTTTt GAAAATCAAT	3900
10	TACCTTGGcA CCTACCAAAT GATTTGAAGC ATGTTAAAAA ATTATCAACA GGTCATACTT	3960
	TAGTAATGGG TCGTAAGACA TTTGAATCGA TTGGTAAACC ACTACCGAAT CGTCGAAATG	4020
15	TTGTACTTAC TTCAGATACA AGTTTCAACG TAGAnGGCGT TGATGTAATT CACTCTATTG	4080
	AAGATATTTA CCAACTACCG GGCCATGTTT TCATATTTGG AGGGCAAACA TTATTTGAAG	4140
	AAATGATTGA TAAAGTGGAC GACATGTATA TTA CTGTTAT TGAAGGTAAA TTCCGTGGTG	4200
20	ATACGTTCTT TCCACCTTAT mCATTkGagr CTGGGAAGTT GCCTCTTCAG TTGAAGGTAA	4260
	ACTAGATGAG AAAAATACAA TTCCACATAC CTTTCTACAT TTAATTCGTA AAAAATAAGG	4320
	GGGAAAACGA CCATGACAAA ACAGATTATA GTAACAGACT CAACATCCGA TTTATCTAAA	4380
25	GAATACTTAG AAGCAAACAA CATTCAATGTA ATTCCTTTAA GTTTAACTAT TGAAGGAGCT	4440
	TCATACGTTG ACCAAGTAGA TATTACATCA GAAGAATTTA TTAATCATAT TGAAAATGAT	4500
	GAAGATGTAA AGACAAGTCA GCCAGCCATA GGTGAATTTA TATCTGCTTA TGAAGAACTA	4560
30	GGAAAAGATG GCTCTGAAAT CATAAGTATT CATCTTTCTT CAGGATTAAG TGGTACATAT	4620
	AACACTGCTT ACCAAGCAAG TCAAATGGTA GATGCTAATG TAACTGTTAT TGATTCAAAA	4680
35	TCTATTTCTT TTGGTTTAGG GTATCAAATA CAACACCTAG TAGAGCTTGT AAAAgAaGGT	4740
	GtCTCAACTT CTGAAATAGT TAAAAAGTTA AATCATTTAA GAGAAAACAT TAAATTATTT	4800
	GTAGTTATAG GGCAATTGAA TCAATTAATT AAAGGTGGCA GAATTAGTAA AACAAAAGGT	4860
40	TTGATTGGTA ATCTTATGAA AATTAAACCA ATTGGTACAC TAGATGATGG TCGCTTAGAG	4920
	CTTGTGcmCA ATGCGAGAAC TCaAAATTck AGTATCCAAT ACTTGAAAAA GGAAATTGCT	4980
	GAATTTATAG GAGATCATGA AATCAAATCC ATTGGTGTCTG CACATGCTAA CGTCATTGAA	5040
45	TATGTTGATA AATTGAAGAA AGTTTTTAAT GAAGCTTTTC ATGTGAATAA TTACGATATA	5100
	AATGTAAC TA CACCAGTTAT TTCTGCACAT ACTGGTCAAG GTGCGATTGG CCTCGTAGTC	5160
	CTTAAGAAGT AAATTTAATC TTTTCAGTGT TAATTACTTC CATTTCATC CTTTATAGAC	5220
50	TAAATTTATA ATTAGATAGA TAGAGGAGGT AATTCATATG ACAAAGAAT ATGCAACATT	5280
	AGCAGGAGGA TGTTTCTGGT GCATGGTTAA ACCATTTACA TCATATCCAG GCATCAAGTC	5340
55		

GAATCAAACC GGCCATGTCG AAGCAGTACA AATTACGTTT GATCCAGAGG TTACTTCCTT 5460  
 TGAAAATATA TTAGACATAT ATTTCAAAAC ATTTGACCCA ACTGATGATC AAGGGCAATT 5520  
 5 TTTTCGATAGA GCGGAAAGCT ATCAACCAGT CATTTTCTAT CATGATGAAC ATCAGAAAAA 5580  
 GGCTGCTGAG TTTAAAAAGC AACCAATTAAG TGAACAAGGT ATTTTCAAGA AACCAGTGAT 5640  
 TACACCTATT AAACCATATA AAAATTTCTA TCCAGCTGAA GACTACCATC AAGATTATTA 5700  
 10 CAAAAAGAAC CCGGTACATT ATTACCAATA TCAACGTGGT TCAGGTAGAA AAGCGTTTAT 5760  
 AGAATCACAT TGGGGGAATC AAAATGCTTA AAAAAGATAA AAGTGAAC TAACAGATATAG 5820  
 AATATATTGT TACACAAGAn AACGGCACTG AACCACCATT TATGAATGAA TATTGGAATC 5880  
 15 ATTTTGCTAA AGGATTTATG TAGATAAAnt TCnGGTAAAC CTTG 5924

## (2) INFORMATION FOR SEQ ID NO: 131:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9280 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

GGCCGTTnAA AATCTCCAAA ATAnAAAAAC CCATCTTGTT CCAATGTTTT AAAATCGCCa 60  
 30 TCCaACACTT GaTCaATAGC TTGCAACAAC GTTGAACGTG TTTTaCCAAA AGCATCaAAC 120  
 GCTCCCACTA AAATCAGTGC TTCAAGTAAC TTTCTCGTTT TGACTCTCTT CGGTATACGT 180  
 35 CTAGCAAAAT CAAAGAAATC TTTAAATTG CCGTTCTGAT AACGTTTCATC AACATCACT 240  
 TTCACACTTT GATAACCAAC ACCTTTAATT GTACCAATTG ATAAATAAAT GCCTTCTTGG 300  
 GAAGGTTTAT AAAACCAATG ACTTTCGTTA ATGTTCCGTG GCAATATAGT GATACCTTGT 360  
 40 TTTTTTGCTT CTTCTATCAT TTGAGCAGTT TTCTTCTCAC TTCCAATAAC ATTACTTAAA 420  
 ATATTTGCGT AAAAATAATT TGGATAATGG ACTTTTAAAA AGCTCATAAT GTATGCAATT 480  
 TTAGAATAGC TGACAGCATG TGCTCTAGGA AAACCATAAT CAGCAAATTT CAGAATCAAA 540  
 45 TCAATATTT GCTTACTAAT GTCTTCGTGA TAACCATTTT GCTTTGCACC TTCTATAAAA 600  
 TGTTGACGCT CACTTTCAAG AACAGCTCTA TTTTTTTTAC TCATTGCTCT TCTTAAAATA 660  
 TCCGCTTCAC CATAACTGAA GTTTGCAAAT GTGCTCGCTA TTTGCATAAT TTGCTCTTGA 720  
 50 TAAATAATAA CACCGTAAGT ATTTTTTAAT ATAGGTTCTA AATGCGGATG TAAATATTGA 780  
 ACTTTGCTTG GATCATGTCT TCTTGTAATG TAAGTTGGAA TTTCTTCCAT TGGACCTGGT 840

55

EP 0 786 519 A2

	ACACTTCTTA CACCGTCAGA CTCTAATTGG AATATGCCAG TCGTATCTCC TTGCGACAAC	960
	AATTCAAACA CTTTTTGATC ATCAAACGGA ATCTTTTCGA TATCAATATT AATACCTAAA	1020
5	TCTTTTTTGA CTTGTGTAA GATTTGATGA ATAATCGATA AGTTTCTCAA CCCTAGAAAA	1080
	TCTATTTTAA ATAACCCAAT ACGTTCGGCT TCAGTCATTG TCCATTGCGT TAATAATCCT	1140
	GATCCCCCTT TCGTTAAAGG GGCATATTCA TATAATGGAT GGTCAATTAAT AATAATTCCT	1200
10	GCCGCATGTG TAGATGTATG TCTTGGTAAA CCTTCTAACT TTTTACAAAT ACTGAACCAG	1260
	CGTTCATGTC GATGGTTTCG ATGTACAAAC TCTTTAAAAT CGTCAATTG ATATGCTTCA	1320
	TCAAGTGTA TTCCTAATTT ATGTGGGATT AACTTGAAA TTTCATTTAA TGTAACCTCA	1380
15	TCAAACCCCA TAATCTCTCC AACATCTCTA GCAACTGCTC TTGCAAGCAG ATGACCGAAA	1440
	GTCACAATTC CAGATACATG TAGCTCGCCA TATTTTTCTT GGACGTACTG AATGACCCTT	1500
	TCTCGGCGTG TATCTTCAAA GTCAATATCA ATATCAGGCA TTGTTACACG TTCTGGGTTT	1560
20	AAAAAACGTT CAAATAATAG ATTGAATTTA ATAGGATCAA TCGTTGTAAT TCCCAATAAA	1620
	TAACTGACCA GTGAGCCAGC TGAAGAACCA CGACCAGGAC CTACCATCAC ATCATTCGTT	1680
25	TTGCGATAAT GGATTAAATC ACTTACTATT AAGAAATAAT CTTCAAAACC CATATTAGTA	1740
	ATAACTTTAT ACTCATATTT CAATCGCTCT AAATAGACGT CATAATTAAG TTCTAATTTT	1800
	TTCAATGTG TAACTAAGAC ACGCCACAAA TATTTTTTAG CTGATTCATC ATTAGGTGTC	1860
30	TCATATTGAG GAAGTAGAGA TTGATGATAT TTAAATTCTG CATCACACTT TTGAGCTATA	1920
	ACATCAACCT GCGTTAAATA TTCTTGGTTA ATATCTAATT GATTAAATTC CTTTTCAGTT	1980
	AAAAAATGTG CACCAAAATC TTCTTGATCA TGAATTAAGT CTAATTTTGT ATTGCTCTCA	2040
35	ATAGCTGCTA ATGCAGAAAT CGTATCGGCA TCTTGACGTG TTTGGTAACA AACATTTTGA	2100
	ATCCAAACAT GTTTTCTACC TTGAATCGAA ATACTAAGGT GGTCCATATA TGTGTCATTA	2160
	TGGGTTTCAA ACACTTGATC AATATCACGA TGTTGATCAC CGACTTTTTT AAAAATGATA	2220
40	ATCATATTGT TAGAAAATCG TTTTAATAAT TCAAACGACA CATGTTCTAA TGCATTCAAT	2280
	TTTATTTCCG ATGATAGTTG ATACAAATCT TTTAATCCAT CATTATTTT AGCTAGAACA	2340
45	ACTGTTTCGA CTGTATTTAA TCCATTTGTC ACATATATTG TCATACCAA AATCGGTTTA	2400
	ATGTTATTTG CTATACATGC ATCATAAAAT TTAGGAAAAC CATACAATAC ATTGGTGTCA	2460
	GTTATGGCAA GTGCATCAAC ATTTTCAGAC ACAGCAAGTC TTACgGCATC TTCTATTTTT	2520
50	AAGCTTGAAT TTAACAAATC ATAAGCCGTA TGAATATTTA AATATGCCAC CATGATTGAA	2580
	TGGCCCTTT CTATTAGTTA AGTTTTGTGC GTAAAGCTGT AGCAAGTTGC TCAAATTCAT	2640

55

	CAATATCATT AATAATCAAT TGCCCTTTAG AACGTAATCG ACATCTGATT TCATTACCTT	2760
	CATCGACTGC AAATACCCAT ATTTTCAAGC CTTTGATGTC AGCAATTGTA TTAACAAACT	2820
5	GAGATGCTTC ATTTGGCTGA ATACCGAATT GCTCCAATAC ATCTTCAGTT ATTTTAACTT	2880
	GGCAGAATCC ATCATCCATA AGTTCGAAAT GTTGTA AAC ATAACCTTGA AACGGCAACA	2940
	TTTTTGGGTC CTTCTCCATC ATTTTATTTA AAAGCGCATT ATGATCAATA TCATGCCCAA	3000
10	TTAACTTTCC AGCAATTTCC ATAGTATGTT CTGAGGTATT GTTAAAAAGG AATCGCCCAG	3060
	TATCACCAGC GATACCAAGA TATAAACGC TCGCGATATC TTTATTAACA ATTGCTTCAT	3120
15	CATTAAATG TGAGATTAAA TCGTAAATGA TTCACTTGT AGATGACGCG TTCGTATTAA	3180
	CTAAATTAAT ATCACCATAC TGATCAACTG CAGGATGATG ATCTATTTTA ATAAGTTTAC	3240
	GACCTGTACT ATAACGTTCA TCGTCAATTC GTGGAGCATT GGCAGTATCA CATACAATTA	3300
20	CAAGCGCATC TTGATATGTT TTATCATCAA TGTATCTAA CTCTCCAATA AAACTTAATG	3360
	ATGATTCCGC TTCACCCACT GCAAATACTT GCTTTTGC GG AAATTTCTGC TGAATATAGT	3420
	ATTTTAAACC AAGTTGTGAA CCATATGCAT CAGGATCTGG TCTAACATGT CTGTGTATAA	3480
25	TAATGTATC GTTGTCTTCG ATACATTTCA TAATTTTATT CAAAGTACTA ATCATTTTCA	3540
	TACTCCCTTT TTTAGAAAAG TTGCTTAATT TAAGCATTAG TCTATATCAA AATATCTAAA	3600
	TTATAAAAAT TGTTACTACC ATATTAACT ATTTGCCCGT TTTAATTATT TAGATATATA	3660
30	TATTTTCATA CTATTTAGTT CAGGGGCCCC AACACAGAGA AATTGGACCC CTAATTTCTA	3720
	CAAACAATGC AAGTTGGGGT GGGGCCCCAA CGTTTGTGCG AAATCTATCT TATGCCTATT	3780
	TTCTCTGCTA AGTTCCTATA CTTCGTCAAA CATTTGGCAT ATCAGGAGAG CGCTCGCTAC	3840
35	TTGTGCTTT TGAATATGCA TGTTCACTTC TATTTTGGCG AAGTTTCTTC CGACGTCTAG	3900
	TATGCAAAG CGCACTGTTA TATGTGATTC AATAGGTACT GTTTAATAT ACACGATATT	3960
40	TAAGTTCTCT ATCATGACAT TACCTTTTTT AAATTACGC ATTTTCATATT GTATTGTTTC	4020
	TTCTATAATA CTTACAAATG CCGCTTTACT TACTGTTCCG TAATGATTGA TTAAAAGTGG	4080
	TGAAACTTCT ACTGTAATTC CATCTTGATT CATTTGTATA TATTTGGCGA TTTGATCGTT	4140
45	AATTGTTTCA CCCATCTGAG GCTGTCTTCC TAAAAGTTGC ATAGACTTTA AAACATCTTG	4200
	TCTATTAATC ACACCCACTG TCTTTTTTATT ACTCGAAACG ACAGGAATCA ATTCAATACC	4260
	TTCCCAAATC ATCATATGCG CACAACCTGC TACTGTACTC ATAGCATTTA CATAAATAGG	4320
50	ATTCGCGTC ATCACTTTAT CTATTTCTGTC GTCGTCCTTT GTATTAATCA TCTCTCGACT	4380
	TGTTACAATA CCTACTAATT TATACGACTC ATTGACTACC GGAAATCTTG TATGGCCAGT	4440

55

ATCTAATGGC GTCATTATAT CTTGAACTAT TAAGATATCT TTTCGTATTT TCTGATTAAA 4560  
 AAGTGCTTTG TTGATAATAT TTGCAACTAG GAATGTATCA TAACTTGATG ATAGAACAGG 4620  
 5 TAAATCATGT TCATTTCGCA AATTAATAAC TTTATTAGAT GGCTTAAATC CACCAGTAAT 4680  
 TAATATAGCC GTACCTCTTT TTAAGCTTC AATCTGCACA TCTTCACGAT TTCCGACAAT 4740  
 CAATAATGTC TTTGGACCAA TATACTTTAA AATATCTTTG AGTTCCATTG CTCCAATTGC 4800  
 10 AAATTTAGAT ACCATCTTAG TGATACCTTT GTTGCCACCT AACACTTGGC CATCAATAAT 4860  
 ATTGACAATT TCATTAAAAG TTAAATGTTT AATTTTCAAT CGATTACGTT TTTCGATTCTG 4920  
 AACCGTACCA ACACGATCTA TCGTTGCGAC CATGCCCAT TATCAGCAT CTTTmATTGc 4980  
 15 ACGATATGCT GTCCCytcCaG ATACGTTTAA AAATTTAGCG ATTTTACGCA CCGAAATTTT 5040  
 AGAGCCTATA GATAACGATT CAATATAATC TAAAATTTGT TCATGTTTTG TCATTCTTTA 5100  
 CCTCTTCTTT TCGAACAGTA TTAACACAT TATAACTTTA TTTTGGATAA AAAGCATTGA 5160  
 20 AGTGAAATGA AATAATGATC GTTtCACCTA TTTTATTTT TGAAAATATA CAACAAACAC 5220  
 AAAGATCACA AAATCTTTAA TTTTAAATGG AAAAATCCAT TATTATTTAT TAGAATGTAA 5280  
 25 GTGAGGAGGG ATGTACTAAT GTATAAAAT ATATTACTTG GTGTAGACAC TCAGTTAAAA 5340  
 AATGAAAAAG CACTAAAAGA AGTGTCTAAA TTAGCTGGCG AAGGTACAGT CGTAACAGTT 5400  
 TTAAACGCAA TCAGCGAACA AGaTGCTCAA GCATCAATTA AAGCAGGTGT TCATTTAAAC 5460  
 30 AAACCTACTG AAGAACGAAG CAAGCGATTG GAAAAACAC GCAAAGCTTT AGAAGATTAT 5520  
 GGTATTGATT ATGACCAAAT AATTGTTCTG GTTAATGCAA AAGAAGAACT ATTAACCAT 5580  
 GCTAATAGCG GTAAATATGA AATTGTTGTT TTAAGTAACC GTAAAGCAGA AGACAAAAAG 5640  
 35 AAATTTGTAC TTGAAGTGT CAGCCACAAA GTAGCAAAAC GTGCGACTAT CCCTGTATTA 5700  
 ATCGTTAAAT AAAATTTTAA TCCAGAATCA CAAATAATCT TTCAATCATG ATGCAGTCTC 5760  
 AAACGACTGA GTAAATACAA GAAACGATTA TGAAGTGGT TCTGGATTTT TTATATCGTA 5820  
 40 GTAAATTTAT AATCAATGTC TAATTGTATA AAATAAAAT TACGAGAGTA GGTCAGAAAT 5880  
 GATAAAGAAC CACTGATGTC CCCCCTCCAC GTCGTAACGT AATCAGTAGA ATATAAAAC 5940  
 ACCCACTAAA AATATGCAGA CGATAACTTC CACATAGATT AGCGAGGTGT TTTTtagTGT 6000  
 45 AAAATCTATA TTCTATTTAA AACTGAACAG ATTCACCTGG TTTTAAATTT TGCACGTCCC 6060  
 CTACATTAAC AGCATCTTTA AATTGTTGTG GATCTTGTTC GATTAATGGG AATGTATCAT 6120  
 50 AATGAATCGG TACAGAAAT TTTGGTTTAA TAAATTCATT AATAGCATAA CTTGCATCAT 6180  
 CAATACCCAT CGTAAATTA TCTCCAATTG GTACAAAACA TACATCAACT GGATGACGTT 6240

55



EP 0 786 519 A2

	TTCAACTTCA AACACGATAC CCATTGGCAT ACCTAAATAA ACTGGgAATA CCATTITTCAT	6360
	GTGTAAAACT TGAACATATGA AATGCTTGAA CAAATTTAAC GCTTCCGAAA TCAAaGTTTG	6420
5	CTTTACCACC AaTATTCATA CCATGAACAT TTTCAACACC GTGATATGAA GAAAGATAGT	6480
	CAGCCATTTC TGCACITCCA ATTACTGTTG CTCCTGTTTT CTTTGCTAGT TCCACAACAT	6540
	CACCAAAATG ATCAAAATGA CCGTGCGTTA AAACGATATA GTCTACCTGC ACTGTTTCAA	6600
10	TATTCAAATC ACACITAGGG TTATTTGAAA TAAACGGATC TACGATAACC TTTTGTGTGT	6660
	TCCCTTCTAA ATAAATCGTT GATTGACCAT GAAATGATAA CTTCATTTGA GCATCCTCCT	6720
	ATCAATTACT ATATAAATTT AGTACCCTTT TGCCACTTAA TTATAACAAA TTCTCAAATT	6780
15	TTAAAAATG AAAATCTAGT TAATGTATTA GCTCGATTTT GAAATCTAAT AATAATTGGC	6840
	ATAAAATGGA AGTAATATTA TGTTGAGGAG TGTTTATAAA ATGACAAAAA TATCAAAAAT	6900
20	AATAGACGAA TTGAACAATC AACAAAGCTGA TGCAGCATGG ATTACAACAC CGTTGAATGT	6960
	ATATTATTTT ACTGGATACC GTAGCGAACC CCATGAAAGA TTATTTGCAT TATTGATTAA	7020
	GAAAGATGGT AAACAAGTAC TATTTTGTCC AAAAATGGAA GTCGAAGAAG TCAAAGCATC	7080
25	ACCTTTCACA GGTGAAATCG TTGGATATTT AGACACTGAA AACCTTTTTT CACTTTATCC	7140
	TCAAACAATC AATAAATTAC TAATTGAAAG CGAGCACTTA ACAGTAGCAC GCCAAAAACA	7200
	ATTAATCTCT GGTTCATATG TCAATTCATT CGGAGATGTT GATTTAACAA TCAAACAATT	7260
30	GAGAAATATT AAATCCGAAG ATGAAATTAG CAAAATACGT AAAGCTGCTG AGTTAGCAGA	7320
	TAAGTGTATC GAAATAGGTG TTTCTTATTT AAAAGAAGGT GTGACTGAAT GTGAAGTAGT	7380
	CAACCATATT GAGCAACTA TCAAACAATA TGGCGTCAAT GAAATGAGTT TTGATACGAT	7440
35	GGTTTTATTT GGAGATCATG CCGCATCACC TCATGGCACA CCAGGAGATC GCAGATTAAA	7500
	AAGCAATGAA TATGTACTAT TTGATTTAGG TGTAATTTAT GAGCATTATT GTAGCGATAT	7560
	GACACGTACT ATTAAATTTG GTGAACCTAG CAAAGAAGCA CAAGAAATTT ATAATATTGT	7620
40	ATTAGAAGCA GAAACATCTG CAATCCAAGC AATTAAACCT GGAATACCAT TAAAAGATAT	7680
	CGATCATATC GCTAGAAATA TTATTTGAGA AAAAGGTTAT GGTGAATATT TCCCTCATCG	7740
45	CTTAGGTCAT GGCCTAGGAT TACAAGAACA TGAATATCAA GATGTTTCAA GTACTAATTC	7800
	TAATTTGTTA GAAGCTGGCA TGGTTATTAC AATCGAACCA GGTATTTATG TACCTGGTGT	7860
	TGCAGGTGTA AGAATTGAAG ATGACATACT TGTCACATAA GAAGGATATG AAGTATTAAC	7920
50	ACATTACGAA AAATAAGGAG TGGGATAAAA ATGAAAAGCT TGTTACAAGC GCATTCTCAT	7980
	TCAGTCAAAC ACTGCCAATA TAACATTGTA GCGCCTAAGA CATAAATTTT TATCCAAGTC	8040

55

5 TGTAATGAAT CAAATCAATA TCATTCATGT TCGATGATT CTTCGCATTG TTTCTAGCTT 8160  
 TAATTTATCA TTATTTAATT TTAATAACCA AGGAGATGAT AACGTCATTG TTTAGTACGC 8220  
 10 TGTAATCCAT TCCCTTTTCA TCAAATTCAA ATTATAATTG TAATGCTTCT TCTACAGATT 8280  
 TATATTCCAT TTCAAATGCC TCTGCAACGC CTTTATTGGT TACGTGACCT TTGTAAGTAT 8340  
 TTAAACCTAA TGATAATGGT TGATTTGATT TAAATGCTTC TCTATACCCT TTATTAGCTA 8400  
 15 GCATGAGCGC ATAAGGTAGC GTAGCATTAT TAAAGCTAA CGTCGAAGTA CGCGGTACTG 8460  
 CACCTGGCAT ATTTGCAACT GCATAATGAA CCACACCATG CTTAATATAT GTAGGATCAT 8520  
 CATGTGTCGT AATTTTATCA GTTGCTTCAA AAATACCGCC TTGATCAATA GCAATGTCAA 8580  
 20 TAATAACTGA CCCATTTTTC ATTTGTTTAA TCATGTCTTC TGTTACAAGT CTTGGCGCTT 8640  
 TAGCACCTGG AATTAAACT GCACCTATTA CTAAATCACT TTGTTTAACT TACAACCTCAA 8700  
 TATTCAACGG ATTTGACATA ATGTATGTA CACGTCCACC GAATAAATCA TCTAATTGTT 8760  
 GTAAACGCTT TGGATTAACTA TCTAAAATCG TAACATCTGC ACCTAGTCCT AGTGCAATTT 8820  
 TAGCTGCATT TGTTCTGCT TGACCACCAC CGATAATAGT TACTTTACCC TTAGGTACTC 8880  
 25 CTGGGACACC ACCTAGTAGA ATTCCCATAC CACCATTAAAG TTTTGTAGG AACTCTGCGC 8940  
 CAACTTGAGC TGACATTCTT CCTGCTACCT CACTCATTGG TGATAACAAT GGTAAAGATC 9000  
 GGTCTGGTAA CTGCACAGTC TCATATGCAA TACTAATTAC TTTTCTATCT ATCAAAGCTT 9060  
 30 GTGTTAATTT TTCTTCATTT GCTAAATGAA gatAaGTGAA TAATACAAGC CCTTCTTTAA 9120  
 AATATGGATA TTCAGATTCA AGTGGTTCTT TAACCTTAAT AACCATATCC ACATCCCAA 9180  
 CTTTGTCTTG TTCAGCAACA ATCTCAGCAC CTGCTTCTTT GTAATCTACA TCTTCAAAGA 9240  
 35 ATGATCCTGA ACCCGcATTT GTTCCACTA AAACAGTATG 9280

(2) INFORMATION FOR SEQ ID NO: 132:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4669 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

CTGATTAATC TCTTGTGTGTC GTGTATTTAC TAATTGAATC GTTGGTGTCT GAACACGTCC 60  
 50 CAGGGATAGC TGTGCATCAT ACTTTGTTGT TAGTGACGCG GTTGCAATTAA TCCCAACAAT 120  
 CCAATCTGCC TCACTTCTCG CTAACGCTGC ATAATACAAA TCGTTATATT GACGACCGTC 180

55

	ACGGATTGGC TTTTGTGTTAC CAACTTTTATC CAAAATCAAT CTTGCAACTA GTTCACCTTC	300
	TCGTCCaGCA TCTGTTGCAA TAATAATATC TTTCACTTTA TTATCTAAAA TTAACGCTTT	360
5	TACTGTTTTA AATTGTTTGC TTGTTTTACC AATAACAACA GTTTTCATAT ATTTAGGTAT	420
	AATTGGAAGG TCTTCTAATC GCCATTCCCTT TAAATTTTTA TCGTATTGTT CAGGTGTCGC	480
	ATTTGTCACT AGATGACCTA ACGCCACGTT GACAATATAT TGGTTATTTT CAAAGTAACC	540
10	ATTACGCTTC TGATTTATTT GTAAAGCATC AGCAATATCT CTTGCGACTG ATGGTTTTTC	600
	AGCTAATATT AAAGATTTCA TAAATTATCC TTTCTCATAC GTTCTTTTAT TTCGAACGTG	660
	CTTCATCTAT TCCACTAATC TTTGATTTAA ATTCAATGAT TGCAAATGAT GTGTTAAATG	720
15	TATTGTAACA TGTTAATATC ACTATTAATC TTCATTTTCTG TTGAAATACT ATATAATAAA	780
	AGTAACAAAA AGTACGGAGG TAATGACATG AGCATAGTTC AGTTATATGA TATTACACAA	840
	ATAAAATCGT TCATTGAACA TTCGAATTAT GAATCAGCAT CATACTTATA TAACTTCCT	900
20	CAACAGTACA ATGAAATAGA TGTATTAATA ACCGATGCGA TTGAATCACC TGGTGTATTT	960
	TCGATTAAAG AAAACGATTG AATCAAAGCA ATCATATTGT CTTTTGCATA CGATAAAAAT	1020
25	AAATTCAAAG TCATAGGCCC TTTCGTGGCT GACAATTATG TATTATCTGT CGATACGTTT	1080
	GAAACGCTAT TTAAAGCAAT GACTTCGAAC CAACCTGACG ATGCCGTCTT TAACTTTTCT	1140
	TTTGAAGAAG GCATTCAACA ATACAAACCA TTAATGAAAG TTATTCAAGC AAGTTATAAC	1200
30	TTCCTGACT ATTACATAGA AGCCCGTACA AGATTAGAAG AAGATATGCA CCAACCAAAT	1260
	ATCATTCCCTT ATCACAAGG GTTTTATCGT GCTTTCAGCA AATTACACAC AACTACATTT	1320
	AAATATCAGG CACAGTCACC ACAAGATATC ATTGATAGTT TAGACGACCA TCATCATTTG	1380
35	TTTTTATTTG TTAGCGAAGG TTTACTTAAA GGTTATTTAT ACCTTGAAAT TGATTCACAA	1440
	CAGTCAATCG CCGAGATTAA ATACTTCAGT TCTCATGTAG ATTACCGTTT GAAAGGTATC	1500
	GCTTTCGAGT TGCTTGCCTA TGCATTGCAA TATGCTTTTG ATAATTTTGA TATTAGAAAA	1560
40	GTTTATTTTA AAATTCGTAA TAAAAATAAT AAATCATCG AACGATTTAA TGGTCTAGGT	1620
	TTCCATATCA ACTATGAGTA CATTAAATTC AAATTCGAAT CACGTAACGT AAAAGATCAA	1680
45	ACAATCCCTG AATAAAACAC CAAGCAAATA CCCTACAGTA CATCATTAGC ATGTATTGTG	1740
	GGTTTTTCTA CTTTTTGTA ATATTGAAAA TTATAAGTAG TTGTTTTTTA CTATTAGGGC	1800
	AGAATGCTTT ACAATAACAT GCAAGTGTC AATTAAGGGGA GCACTTGCAT AAATAGTATA	1860
50	GGAGAGTGAG TAGTCTTGCA ATTTCTTGAT TTCTTAATCG CACTTTTACC TGCTTTATTC	1920
	TGGGGAAGTG TCGTTCTTAT TAATGTGTTT GTCGGCGGTG GACCTTACAA CCAAATTCGT	1980

55

	TTCAATAATC	CTACTGTAAT	TATTGTCGGT	CTTATTTCTG	GTGCATTATG	GGCGTTTGGA	2100
	CAAGCGAATC	AGCTTAAATC	TATTAGTTTA	ATCGGTGTAT	CAAATACTAT	GCCAGTTTCT	2160
5	ACAGGTATGC	AATTAGTTGG	TACAACATTA	TTCAGCGTTA	TCTTTT TAGG	TGAATGGTCT	2220
	TCAATGACTC	AAATTATCTT	TGGTTTAATC	GCCATGATAT	TATTAGTTAC	TGGTGTAGCA	2280
	CTTACTTCAC	TTAAAGCTAA	AAATGAACGT	CAATCAGATA	ATCCTGAATT	TAAAAAAGCA	2340
10	ATGGGTATTT	TAATTGTATC	TACAGTTGGA	TATGTAGGTT	TCGTTGTACT	TGGTGACATC	2400
	TTTGGTGTG	GTGGAACTGA	TGCATTGTTT	TTCCAATCTG	TCGGTATGGC	AATTGGTGGC	2460
	TTTATCCTAT	CCATGAATCA	TAAAACATCA	CTTAAATCAA	CAGCACTTAA	TCTATTGcCA	2520
15	GGTGTGATTT	GGGGAATTGG	TAACTTGTTT	ATGTTCTATT	CTCAACCAAA	AGTTGGTGTG	2580
	GCTACAAGTT	TCTCATTATC	ACAGTTACTT	GTTATCGTTT	CAACCTTAGG	CGGTATTTTC	2640
20	ATTTTAGGAG	AAAGAAAAGA	TCGTCGTCAG	ATGACGGGTA	TTTGGGCAGG	TATTATTATT	2700
	ATCGTGATAG	CTGCTATAAT	TCTAGGTAAT	TTGAAATAGA	AAGTTAAATA	CTCATGTAAC	2760
	GTAAAAATGT	AATCACTTCT	GAAAATAACC	ATTCAC TTAT	AGAATGATTA	AAATTAATTT	2820
25	TCGGGAATTT	TACGTTGAAT	GTTCCCTCTAT	ATGTCCTAGG	AAATACGTGG	CTCTAAAAAC	2880
	AAAACGCAAT	AACACATCAT	GACATTAATC	ATGCGTTTTA	AGACTTTAAA	ATTAGCGATA	2940
	CTTTTAAAT	CTTGATGATA	TTCATATATC	AAGTATGCGC	CATACATATG	AAGTGGATAG	3000
30	CTGCATAACG	CACTGCATTA	TCAACTTGAA	TGTATGAGTT	GAACAACTAT	GTCAATAAATA	3060
	AAAGCCCCCT	TTTCACAATA	TACATTTACA	TATTGTGGTA	AAGGGGGCTC	TCATTTTCTA	3120
	CGAATACTAA	AATGGATTTT	ATTTTCAAAT	GTGTAAACTA	GACAAACACT	GCCTGATACA	3180
35	CGTACAAAAT	AATGATACTA	ATAATGATTG	TCAAATTGGT	CGTCATACCT	ATAAATGGCA	3240
	GTGTTTCGATA	TTTAAACTGA	ATACCATAAG	AAATAATTGC	AACACcTACC	GGGAACATCC	3300
40	AAGTGACCAA	CAATGTCGTC	TTAATCATAT	CATCTGATAC	TGGTAACAAC	ACATATACTA	3360
	ACAATCCCGC	AACTAATGCT	AATCCATAAT	GCAAACATAA	ATATTTAATA	GTAGCAGGTA	3420
	TATACTTTCT	TTCCAGAGTA	AAATTCAACA	TGACACCTAG	CAAAATCATT	GATAACGGCA	3480
45	TATTTGCATG	GGAAAGTATG	CTAAAGAAAT	CGATTGCCAC	ATGTGGTAAA	TGGATGTGAC	3540
	TTATATTCAA	TATAAACATT	ACAATGTATG	TAACGAGTGG	CACTGATTGT	AATAATTTCT	3600
	TACCTAAATA	TTTAAAATCG	AATTGATCAC	TACCTTCACT	AAAGTAGCTA	CCTACAAAGT	3660
50	AAGTAATTCC	AAACATCACA	AAGGCACCAC	CTATATCAGC	CATAACAAAA	TAAATAAGTC	3720
	CCGTTTTAGG	CCATATCACT	TCAATTAGTG	GATATGCAAA	CAATCCAATA	TTCATAGCAC	3780

55

CAATCATT TTT CGCCACAATA CCATATATAA TCATTAAAAAT TGGTAAAATG GAGAATGACA 3900  
 ATTTTAATTC TGCAC TGT TTT AAATTCACAA TAACTAAAGA TGGGAGTGTG ACATTAAGAA 3960  
 5 CTAATGTAGC AATGACTTGA CTATCTGTTG CTTTTATAAA ATTAATGCGC TTCAAAAAGT 4020  
 AACCAAGCGC AATTAATAAA ATAATCATAG TAAATTGTTT TGTCAC TGT T ATCCCTTCTT 4080  
 TCAATAATCT TCATAATTTA TAACTTTAAC ATACTCCACA GATATTTTAG AAGTCTACTG 4140  
 10 TTTCATGCTA TAATCTACAT TAAATGCACT TAATTATATT TCAAAGGAGT GTTATAGTAT 4200  
 GTCTTTAGAA AACCAACTAG CCGAACTTAA ATATGATTAT GTTCGTCTTC AAGGTGACAT 4260  
 AGAAAAACGG GAATCTTTGA ATTTAGATAC TTCCGCACTT GTTCGTCAAC TTAAAGATAT 4320  
 15 TGAAAATGAA ATTAGAAACG TTCGTGCTCA AATGCAAGAT TAATAATCTA TCATTCAAGC 4380  
 AATAAATGCT TTTTGTTACA TAAATTTGAC TAGCATTGCT CTGAATACGT TATATTGATG 4440  
 AATTGCTTCA TTTTTCGCTC AATTACATCT AGAATCACAA GATGTTGTCG TGTTATGATT 4500  
 20 TAGTGTTTCA TTAACAACAT ACACGCATAT CTATCCCAAC ACTGCTATTT ATGTTTTCTA 4560  
 CGCTGnTGTA CTACATGAAC CCTTTGAAAC GGAGAGGAAG TTATCATATG CAATTTTAnC 4620  
 25 TGATTTTACT AGCAATACTT TAACnAATTG nTAGTTTAAT AGAATTTTA 4669

(2) INFORMATION FOR SEQ ID NO: 133:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 2785 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

TTTGcACCCA TCTGaTACAA TGCACCATGC GGTTTAACAT GATTAATTTT AACTTGATGA 60  
 ATGCGACAAA ACCCTTGTA TGCACCTAAT TGATAAATCA TCAAATTATA AATCTCGTCG 120  
 40 TTAGAGATAT CTATATTTTCG TCTGCCAAAG CCTTTCAAAT CAGGTAAACC AGGATGTGCA 180  
 CCTACTGCAA CATTATGTGC TTTGGCAAGT TTTACCGTTT CATTCAATTAC ATTTTCATCA 240  
 CCAGCGTGAA AACCACAAGC AACATTCGCA CTTGTAATTA ACGGAATAAT TTGATGATCA 300  
 45 CCACCAAAGG AATAATTTCC AAATGCTTCG CCTAAATCAC AATTCAAATC AACTCGCATT 360  
 ATAATTCCAC CCCTTTAACA ATTTGATGTT TTTCTAAAAA TTTAATATCA ACATCTTTTG 420  
 50 CATCTCCATC ACGATATAGT GGATAATTTA AAAGTGCATA TAAAAAATCG GCAGTTGTAG 480  
 AAAATCCATC TATCACCATT TCATCTAAGG TGACTTTCAA CTTATCAATT GCTGAAGCTC 540

55

## EP 0 786 519 A2

	AACCGTGATA TAGTAAAGAA TCGACTCGCA CATTAAAGCC TTGAGGTAAA TGTAACGCTG	660
	TCACTTTACC TGGTGTGGT TGAAATTTCT TTTCaGGATT TTCGGCATT ATTCTCGCTT	720
5	CTATCACATG ACCATTAAAT TGAATATCGC TTTGTGAAAA AGGTAAATGA TTATGTTCCA	780
	ATAAATACAG TTGTGCTGCA ACCAAATCAC GTTCTGCTCG CATCTCTGTA ACAGTATGTT	840
	CAACTTGAT TCGAGCATT ATTTCAATAA AGTAATGTGC GGTATCAGTT ACTAAAAATT	900
10	CAATCGTACC TGCATTCTA TAATTTGCTG CACGTGCAAC TTAAACAGCA TCGTTACATA	960
	TTTGTGTGCG TCTTTCTTCA GTTAATGCTG CACAAGGAGA TTCTTCGATT AATTTTGTAT	1020
15	TTTTACGTTG TACAGAACAA TCACGTTCCC CTAAATGTAC ATAATTATCC TGCCCATCTC	1080
	CCaTAACTTG AACTTCAACA TGTTTTGcAA CAGGTATAAA AGCCTCAACA TAAACACGAT	1140
	CATCATCAAA GTATTTTTTT CCTTCACTTT TAGCTTCTTT AAATGCCTTT TCTAAATCTT	1200
20	CAGCTTTCTT TACAATACGT ATACCTTTAC CACCACCGCC ACTGGCAGCT TTGATAACAA	1260
	CTGGATAACC GATGTCTTTG GCAAGATTCT CAATTCAGA CACATGATTC ACAGCACCAT	1320
	TTGATCCTGG AATCACAGGA ACACCTGCAT GATGAAGTGT TTGTCTTGCT GTTATTTTAT	1380
25	CCCCCATCAT TTCCATCGTT TTTTtagTAG GCCCTATAAA CGCTATGCCT TGTTCTCTAA	1440
	CGGTTTGAGC AAATTTTGTT GATTCTGATA AAAAGCCATA TCCTGGGTGA ATTGCATTAG	1500
	CACCAGTGAT TTGTGCAGCA GATATGATGC GGTCAATATT TAAATAACTA TCTAAAgCAT	1560
30	TArCwTCCCC AATACATATA GCTTGATCTG CTAAATGTAC ATGCAAGCTT TGCTCGTCCC	1620
	CTTTTGATA AACTGCTACA GTTTCAATCC CATATCTCT GCAAGCTCTT ATAATCCTTA	1680
	CAGCAATTC ACCTCTGTTC GCAATTAAAC AACGAAGCAT TTACTTACCC CCTTTACTTA	1740
35	ATACGTACCA AACTTGGTC GTATTCAACA TTTGTGCCAT GATCAGCTAC TATTTAGTA	1800
	ATTCTCCAG CAACATCTGT TGTTACCTCG TTTAATACTT TCATCGCTTC AACATATCCT	1860
40	ATAATATCTC CCTTGTTAAC TTTGTCACCG ACATTCACAA TTGGTTCAGT TAATTCTTTA	1920
	CTATCTTGTA AAAAGAATGT ACCTATCATT GGTGATTTAA TGTCATGATA ATCATTGTGTC	1980
	GAAACATCGG AGTTATCATT CGCTTTTGAA GCTGTCAAAT CATTATTGTT CATACTTTGA	2040
45	TTGATTGAT TACTGTGTGC AGCCAAATGA TTCGAGTCAG TGAAGTCAAT TTCTATTTCa	2100
	TCCTCAAAAT TTTTATATTT AAATTTCTTA ACATCATTTT CCTTCACTAA TTTGATTATT	2160
	TGTTTCGATTT nTTCAATATT CATTTTACAA ATCCCCTTTT AAAATTGTTG CTAATTTTTT	2220
50	CGAAGTATGT CGCAAGCTAG ATGTATCAAA AATTGGAGTC TTTTGATGAC TCTTAAGAAT	2280
	TTCATTAAAC AGAGACATTT GTTCCCGATT CTTATCTACA GCTTCTTGGA ATGATATCCA	2340
55		

EP 0 786 519 A2

TACAGTTGCA ATTTTGGTAT AACCACCTAT CGTTTGTTTA TCATTAAGCA GAATAATAGG 2460  
 TTGACCATCA TTTGGTACCT GAACACTACC AAGAGCAACC GGTTCAGAAA TGATATCTGC 2520  
 5 TTGATTAACT GGTGCAACGC TGTCACCTTC CAAACGATAG CCCATACGGT CTGATTGTTC 2580  
 AGTAATTAAA TATGGATGAT TTACAATTTT CGCTCTAGCC TCTTCAGAAA ATGCCTCGAA 2640  
 TTGAGGTCCT TGAAGAATGT GTATAATATT ATTTTCTGGC AATAAATCGT CCTGTAAATG 2700  
 10 AATCGTCTTT CCAATGTTTT CTTTAAAGTC ATTATTTATT TTCACTGTTA TTACATCATC 2760  
 AGCTAATAAC TTTCTACCTT TGAAT 2785

(2) INFORMATION FOR SEQ ID NO: 134:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1010 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

25 AATGGAAACG GTTGAAACAG CAATTATTAC TATTTCTATG GGTGAAGGTA TTTCAGAGAT 60  
 ATTTAAATCA ATGGGTGCCA CACATATCAT TAGTGGTGGA CAAACGATGA ATCCTTCTAC 120  
 AGAAGATATC GTTAAAGTCA TTGAACAATC AAAATGTAAA CGTGCAATTA TTTTACCGAA 180  
 30 TAATAAAAAT ATCTTAATGG CAAGTGAACA AGCAGCGAGT ATTGTTGATG CAGAAGCTGT 240  
 TGTTATTCCA ACGAAATCTA TTCCTCAAGG TATAAGCGCA CTATTCCAAT ATGATGTGGA 300  
 CGCAACACTT GAAGAAATA AAGCGCAAAT GGCTGATTCA GTAAATAACG TTAAATCTGG 360  
 35 TTCATTAACG TACGCTGTTC GTGATACGAA AATTGATGGC GTTGAGATTA AAAAAGACGC 420  
 GTTTATGGGC TTGATTGAAG ATAAGATTGT AAGCAGCCAA AGTGATCAAT TAACAACGGT 480  
 TACTGAGTTG TTAAATGAGA TGTTAGCAGA AGATAGTGAA ATATTGACTG TGATTATTGG 540  
 40 TCAAGATGCA GAGCAAGCAG TTACAGATAA CATGATAAAC TGGATCGAAG AGCAATATCC 600  
 AGATGTAGAA GTGGAAGTTC ATGAAGGTGG ACAACCAATT TATCAATATT TCTTTTCAGT 660  
 AGAATAAAAA TTTAAATATA AAAACTACCA ATGATAAATC ATCAGTTGGT AGTTTTTTAT 720  
 45 TTTGCTATTT TAGTGATATT GCGGGTTAAA AGTATCGTTC TCGAGTTGCT AACAATGTCA 780  
 TGTTCAACTT AGTCATGATA AAATAAATAA CATACTAAAT GATACGTAAA ATCAAATAAA 840  
 50 ACATAGGTGA TTTATTTTGG CTAAAGTAAA CTTAATAGAA AGTCCATATT CTCTTTTACA 900  
 ATTAAGAGGT ATAGGTCCTA AGAAAATAGA AGTATTGCAA CAACTAAATA TTCATACAGT 960

55

## (2) INFORMATION FOR SEQ ID NO: 135:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1540 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

```

TG TAGTTGAA CATGAACAAC AAAAGAAAGA AAAGACAAAA AAGCAATACA AGCCATTTTG      60
GATTGT CATG AGTTTTATAA TACTTATAGT TGTACTATTA CTCCGGGCAC CTTCAAGTCT      120
GCCGATAATG GCTAAGGCAG TACTAGCTAT TTWAGCTTTT GCAGTTATTA TGTGGGTAAC      180
GGAAGCTGTA TCATATCCGG TGTCAGCAAC TTTAATTATT GGCTTAATGA TATTACTTTT      240
AGGATTTAGC CCTGTTCAAA ATTTAGGGGA GAAGCTAGGT AATCCGAAAA GTGGCAGTGC      300
TATTTTAGCT GGAAGTGACC TTCTAGGAAC TAATCATGCA TTATCATTAG CGTTTAGTGG      360
ATTTGCAACT TCAGCTGTAG CTCTCGTTGC AGCTGCATTA TTTTGGCTG CTGCTATGCA      420
AGAAACGAAT TTGCATAAAA GACTAGCTCT TTTAGTGTTA TCAATTGTTG GTAATAAAAC      480
TAGAAATATA GTTATTGGAG CAATTATCGT TTCAATTGTA CTTGCATTTT TCGTTCCTTC      540
TGCAACAGCT AGAGCAGGGG CAGTTGTACC AATCTTGCTG GGTATGATTG CGGCATTAA      600
AGTTTCCAAA GATAGCAAGT TAGCGTCTTT ATTAATAATT ACTTCAGTAC AAGCTGTGTC      660
AATTTGGAAT ATTGGTATCA AAACGGCGGC AGCACAAAAT ATCGTAGCGA TTAATTTTAT      720
AAACCATCAA TTAGGATTTG ATGTTTCATG GGGCGAGTGG TTCTTATATG CAGCGCCTTG      780
GTCCATAGTT ATGTCCGTAG CTTTATATTT CATCATGATT AAAGTGATGC CTCCAGAAAT      840
TAATACAATA GAAGGTGGTA AAGATTTAAT AAAAGAAGAA TTGCATAAAC TTGGCCCCGT      900
TAGCCACAGT GAATGGCGTT TAATTGTTAT ATCGATGTTA TTATTACTGT TTTGGTCAAC      960
TGAAAAAGTA TTACATCCGA TTGACTCTGC ATCCATTACT ATTATTGCTT TAGGTGTTAT     1020
GTTAATGCCG AAAATTGGTG TCATGACATG GAAACATGTT GAAAATAAAA TACCATGGGG     1080
AACAAATTATC GTGTTTGGTG TAGGTATTTT ACTAGGTAAC GTTCTTTTGA AAACAGGTGC     1140
AGCTCAATGG TTAAGTGATC AAACTTTGGG TGTTTTAGGT TTAACACATT TACCTATTAT     1200
CGCGACAATT GCACTTATCA CGCTTTTAA TATATTGATT CATTTGGGCT TTGCGAGTGC     1260
AACAAAGTTTA TCATCAGCGT TAATACCTGT TTTTATTTTG CTAACCTCTA CGTTACACTT     1320
AGGAGACCAG TCTATAGGAT TTGTTTTAAT TCAACAATTT GTTATTAGTT TTGGTTTCTT     1380

```



AGATTTCCTTG AAGGCAGGTA TACCATTGAC AATTGTAGGG aATAtctAgT GaTAGTTTTT 1500  
 AGCATGACTT ATTGGAAATG GGTAAGGTTG CnTTAATTAA 1540

5 (2) INFORMATION FOR SEQ ID NO: 136:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11823 base pairs

(B) TYPE: nucleic acid

10 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

15 ACTTCTCACA ATAAGAAATA TGAAATTGTT ATGTGTTAGT TGAGATTCAG TGATGAATTA 60  
 CTTTTATCAT TTAAATGTT GTTATCATTG TCATGCGTTA CCAAATCGCT TACGTATACA 120  
 CGATTCCCAA TCTTAACATA GACGATTTGT ATATCAGAAT TTTCTGATTA CTAACAGTTT 180  
 20 ACCTAAGTTT AAATATCTGT TCAATGATT TCAGTTATTT TTAAAAGAAA AATCGTAATG 240  
 CTGCCATGAT AACAATCCCA CTAATAATTG TAATAGTTAA AtACGCGTGA TTATAGATAA 300  
 25 AATAACCGTC GGAATGAGCG CGATAATGTA AGGGATGTTT AATGTATACC CCTCACCATG 360  
 AGGCGTCTGT TGAATAATGC TGTCAATGAC AAGTGCCGTA AATAGTGTGA TTGGGATAAA 420  
 TGATAGCCAT CGAACCACGA CATCAGGCAA TTGCACTTTT GAAATCATGA TAAAAGGTAT 480  
 30 AATTCGAATT AATAGCGTTA CGATACCACA CAATAAAATA AGTATTAACA TGTTCATATG 540  
 AGTTATCATT GTTCCATCAT CACTCCTAAC GCTGCTGAAA TTGTGGCTGC AATTAATATT 600  
 GCTAGATATG AAGGCATAAA CATACTTAGC GATAACATCA TTACTATGAC GGCAATAATG 660  
 35 AGTACTATGT AAATCTTAA TCGCGATTTA GTAATTGATT CAAATGCGC AATGGCCAAA 720  
 AAGATAAACA TAGCCGTGAT AGCAAAATCT AACCCTAGCG TTTGCGGATT TGAGATATAT 780  
 TCGCCAAATA AAGCCCCAGC TACACATGAA ATTGCCCAA ATAAATATGC TGTGATGTTA 840  
 40 AGACCATGCA TCCAACGATC ATTGATAGCT TCTCCTTTTA AATAAGGTGT AATGGCGACG 900  
 CCAAACGTTT CGTCAGTTAC TAATGAACCT AATCCAACAC GGTTCACAAA CCCATATGTC 960  
 TTGAAGTTTG GTGCAAGCGA CATACTTAAA AGGAACATTC TTGAATTTAC GATAAATACA 1020  
 45 GTTAGTACAA TCGTGATAT AGGTGTACCT GCTATAAACA ACGCGCACAT AATAAATTGC 1080  
 GCAGcaCCGG CATATATAAC AAGACATAAC AAGACAATTT CTAAAATACT AAAGTTTTGA 1140  
 50 GACGAAGCCA CAATACCAAA TGAAATACCA ACACCGGCAT AACCCAATAA TGTTGGGATA 1200  
 CACTCTTGCA CGCCTTGTCT AAAACTTAAA TGTGTTGTCA TCTCAATTAC CTCCTTTGCC 1260

55

	TAAGCAATAA CATTAGACAT CAGTTTGTCT GAGGTTAGAC ATTCCGGAGT CTTTAGTCAG	1380
	CTTCATATTA ACTTTTTATT TTTGAGAATT TTCAATTTTT TATTTAAGAC TACCTCCATA	1440
5	TTTTCTATGG aTTTGTAGTT GTTTTTAAGT ATCAATTTTA TAAATTTTAA TATCTGATGA	1500
	TGAGTCTGGG aTATTGaTTC ATGTACCACT CCCTTaTaAT CATCCCCTCC CCCTaCCCTA	1560
	CTCCATCGAT ATAACTCATA CTACATATCA ACGAAATCAG TATTTTATCG CTCCTTTTCC	1620
10	TATATTAGTG ATGCTCAAAC TTGTTACGTT TTAGATTGTT TTAGTTCATC ATAATTATCC	1680
	CGTATTGTTG CTATAATGAA ATGCGTTCAC CCCATTAAAC CACAAACTTA ATTTATTGTT	1740
	GTTATGTGCA TTGGCTCACT ATTATATTTT TACAGCACAA AAAAAGTGGC GACAGTTCGT	1800
15	CACCACTTTT TAAAATATTA TTTAAAGTAT CTTGCCCTTG CTTTAAGTAT ACGTAGATAT	1860
	ATACTTTTTA AAGCTTGTAG CTAAAGCCTT TATTTAACTG GTTTTGAAAT TTGTGTTTTA	1920
	CCACCCATAA ATGGTACTAA TGCTTCTGGA ATTGTTACTG TTCCATCTTC ATTTTGGTAA	1980
20	TTTTCAACAA TAGCAGCAAA TGTACGTCCA ACTGCTAAAC CACTACCATT TAATGTATGT	2040
	GCTAATTCTG GTTTAGCTGC TTTGTCACGC TTGAAGCGGA TGTTAGCAGC ACGCGCTTGG	2100
25	AAATCCGTAC AGTTTGAGCA TGAACAAAT TCTTTATAAT CATTGTAGCT TGGTAACCAA	2160
	ACTTCTAAAT CATATGTTTT GCTTGCACTA AATCCAATAT CACCTGTACA TAAAATAACA	2220
	CGACGGTATG GTAAACCTAA CTCTTCTAGA ATTGCTTCTG CGTTTGTGTG CATTTCTTCT	2280
30	AAAGCATTCC ATGAATCTTC AGGTTGTTCA AAACGTACCA TTTCCACTTT ATCGAATTGA	2340
	TGTAAACGAA TTAATCCTCT TGTATCTCTA CCTGCTGATC CTGCTTCACT ACGGAAACAT	2400
	GCAGATTGAC CAGTGAATTT TTCAGGAAGT ACACCTGGTT GAATAATTTC ATTACGGTAG	2460
35	AAATTCGTTA ATGGTACTTC AGCAGTTGGA ATTGTATATA ATCCTTCTTT TTCTACTTTA	2520
	AATAAATCTT CTTCAAATTT AGGTAATTGA CCTGTACCAT ACATTGTATC TGCGTTCACA	2580
	AGCTGTGGTA CCATCAATTC TGTATAACCA TGTGTGTTG TATGTTTTGT AATCATATAG	2640
40	TTCAATTAAAG CACGCTCTAA TTGCGCACCT TCATTTGTTA AATATACAAA ACGCGCACCT	2700
	GAAACTTTTG CTGCACGATC AAAATCAGCC ATTTTCAATT CTTCTACAAT ATCCCAATGT	2760
45	GCTTTGGGTT CAAATGAAAA CTCaCGTGGT GTACCCCACT TTTTAACTTC AACGTTATCT	2820
	TCATCAGATT CACCTTGAGG TACATCATCA CTTATTAAAT TTGGAATACG ACAAAGGATA	2880
	CCTGTCAATTT TATTATCAAT TTCATTAAAT TGAATATCTT TTTCTTTAAT ATCGTCACCT	2940
50	AATGTGCGCA TTTCAGCAAT CACATCATCA GCATTTTCTT TATTACGTTT TTTTAATGCG	3000
	ATTTCTTCGC TTACTTTATT ACGACGTGCT TTCATTTCTT CTGTTGCACT AATTAATTTA	3060

55

	TCAATTTTGC TCTTAACTGT GTCAGGCTCA TTTCTGAATA ATCTAATGTC TAACATTAAC	3180
	CTTCATCCTT TCCCAAATAA TTATCATTTA TTATGGAATG ACGTACGTCT TTATTTTTTA	3240
5	GAAAATAAAA AAAGACCACA TCCCTACAAG GGACGTGGTC TACGCGTTGC CACCCTATTT	3300
	AACAAITTTAA GTTATAAAGA TACACTAAAC CTAAATTGCA CTTCACTAAA ATAACGGTTA	3360
	TCACCGATTG TTCTTTTAAA TTAAGTAGGT AGATTTCATAT ATATGTTGAT TCTTGTTTAC	3420
10	ACTAACCACA AGCTCTCTGA TATCGAACAC TATATATTAC TTGTCCTACG AACAAATGTCT	3480
	TATTAAGTTA TTTTAAATAT AGCAAATAT ATTTGCTTTT TCAAGTAACG ATTTCAAACA	3540
	TCACTCATGT CGATTTAGTG ACATGCAGTC GTTTGATAAA TTGATTGCTT TAAATACTGT	3600
15	GCAACCGCTT CAATATCTTT ATGAAATTGA CGATCATGTG TAATGGATGG CACGATACTT	3660
	CGAAACTCAT CATACTTGCG ACGTGTTTTT GGTGATAATC CTTCAACACC TTTTAACTCT	3720
20	GCTGCTTGTA ATGCAATAAC ACATTTCGATT GCCAGCACAC GTCTTGCAAT TTCAATAATT	3780
	TGATAACCAT GTCTAGCAGC TGTAGTTCCC ATAGATACGT GATCTTCTTG GTTCGCAGAT	3840
	GAAGTGATAG AATCAACACT CGCTGGATGC GCTAAAGTTT TATTTTCAGA AACGAGACTT	3900
25	GCAGCAGCAT ATTGCCATAAT CATCGCGCCA CTTTGCAATC CTGGCTCTGG ACTAAGAAAT	3960
	GCTGGTAAAT CACCATTTAA TTGAGGATTT ACTAGTCGCT CTAGACGACG TTCCGATACG	4020
	TTTGCTAATT CACTTACACC TAATTTAAGA TGATCTAATG CAAAAGCAAT AGGTTGTCCA	4080
30	TGGAAGTTAC CACCTGAAAT AACAAACGTT TCATTGCTT CCTCAAATAT AAGTGGATTA	4140
	TCATTAGCCG CATTCAATTC AAATTCTAAT TGCTGTTTAA CATAATTGAA TACTTGAAAA	4200
	CTCGCGCCAT GGATTTGTGG TATACAACGC AACGTATATG CATCTGTAC ACGTATTTCT	4260
35	GAITGTCGCG TCGTTAATGT TGATCCTTCT AACCAATCAC GCATACGCGC TGCCACATTA	4320
	ATCTGTTCTT GAAAATTACG AACTGCGTGC ACATCATGTC GATATGCATC TATAATGCCA	4380
	TTAAGAGACT GATGCGTTAA TGCAGCAATC CATTGAGATT GGTAACCTAA ATCTTCTGCT	4440
40	TCTATATAAC TAATGACACC TTGAGCTGTC ATAGCTTGCG TACCATTAAAT CAATGCTAAA	4500
	CCTTCTTTAG CCTGAAGGTT CAAAGGTTGT CTATTTAATT CTCTTAATAC ATCGTCACTA	4560
45	TCCTTTTCTT CCCCTCTGTA CAATACTTTC CCTTCACCAA TTAATGCTAA TGCTAAATGT	4620
	GATAATGGCG CTAATCTCC TGATGCACCG AGAGAGCCTT GCTGTGGGAT TATCGGTATA	4680
	ATACGTTTAT TTATAAAAAA TTGTAATTGT CTCCTAATT CTAAAGTGGC ACCTGAATGA	4740
50	CCTTTTAAATA ATGTATTCAA TCGTAAAATC ATCATGACTA ATGCTACTTC TTTTGAAAAAT	4800
	GGCTCACCTA GTCCACAGGC ATGTGAGCGT ATCAGATTCA CTGTGAATTC ATTATATTGC	4860

55

	TCCTCATTTT CAATAATACG TTCAACTACC GCTCTACTTT TTTTGACACG TTCTAACGCA	4980
	TCATCAATAA TTTCAATCTT TGATTGTTGT TGTA AAAATG ATTTAATATC CTCAATTGTT	5040
5	AGTGTTCAC CATCTAAATA TAAAGTCATA TATGTTACCC CCTTGTTTAT ATTAAGTAAC	5100
	CCATCCTTCT TGAAGTATAC GTTTTCATTT TTATTGAAAC AATGGTTTTA CGTACATTTA	5160
	TAACCTATTA TCAGAGCACT ATTGTAGTGC GTTAAAGGAT ATTAAGATTG TTGTAAGCAT	5220
10	ATTTAATAAT TTATCTATTG ACGAATTGCA TATACAGGTA TAGTATTTTC TATTGTATTT	5280
	AACGACAAAT AATAATGAAT TCAGAAATTT ATAATACATT TTGTTAAAAG TTAATATATA	5340
	TTTTTAAAT TGAATAAATT CGGAAAAGGC TTTTACATGG GAGGTTATAT CACTATGGAA	5400
15	ACGTAAAT CTATTAACAT TCCTAAGCGT AAAGAAGATT CACATAAAGG TGATTATGGC	5460
	AAAATTTTAT TAATTGGTGG ATCTGCTAAC TTAGGTGGTG CCATTATGTT AGCGGCTCGT	5520
	GCATGTGTAT TTAGCGGTAG TGGTTTAATC ACTGTAGCTA CACATCCAAC AAATCATTCA	5580
20	GCATTACATT CTCGTTGCCC AGAAGCGATG GTTATTGATA TTAATGATAC GAAAATGTTG	5640
	ACGAAAATGA TTGAAATGAC TGACAGTATA CTAATTGGTC CAGGTCTTGG CGTTGATTTT	5700
	AAAGGAAATA ATGCCATTAC ATTCTACTA CAAAATATAC AACC GCATCA AAATTTAATC	5760
25	GTAGACGGCG ATGCGATTAC AATCTTTAGT AACTGAAAC CGCAATTACC TACATGTCGT	5820
	GTGATCTTTA CACCACACCT CAAAGAATGG GAACGATTAA GTGGTATTCC TATTGAGGAA	5880
30	CAGACATATG AGCGTAATCG TGAAGCAGTT GATCGTTTAG GTGCAACTGT TGTACTTAAA	5940
	AAACATGGTA CTGAAATTTT CTTTAAAGAT GAAGACTTTA AATTGACAAT CGGTAGCCCA	6000
	GCAATGGCGA CTGGTGGTAT GGGCGATACA CTTGCTGGTA TGATTACAAG CTTTGTGGT	6060
35	CAATTTGATA ACTTAAAGA AGCGTTATG AGTGCCACAT ATACACATAG TTTTATTGGC	6120
	GAAAACCTTG CAAAAGATAT GTATGTGGTG CCACCATCAA GACTTATCAA TGAAATACCT	6180
	TACGCAATGA AACAATTAGA AAGTTAGTCA TTAATAATCA TTGAATATAG TAAAGCATT	6240
40	CTTTCTAGCA TAAAAATAAG ACTCCCTTAC ATATAGGGAA GTCTTATTTT TTATTATTCT	6300
	TCATCTGATG ATTGTTGTAT ATCTTCTTCA ACACGATCCA TGAAATCTTG TCTTACTTCA	6360
	ATACGTCCAT CTTTCATCATT TTCTTCTGAA TCAATCACTT CAGTATGAAT TGCATTTCT	6420
45	GGTGTTCAT CATTTaCAAC CGCTTCACGT TGTTGTTTCA TACCATCTTC AGATACAGTT	6480
	GAAGTAGATT GCTCATCTTC ATTCGTTTCA TCTTCTGCAT CTTCTTTTAC TTTAGCAACC	6540
50	GTTGAAACAA ATTGATCATC ACCTAAGCGA ATTAAGCGAA CACCTTGTGC TGCACGACCA	6600
	TTTTGAGAAA TATCTGCAAC ATCTAGTCGA ATAATGACAC CTGCATTAGT AACCAATCATT	6660

55

	GTAGCTGTTT TAATACCTTT ACCACCACGA TTTGATAAGC GATAGTCATT AACTGGCGTA	6780
	CGTTTACCAT AACCATTTTC AGTAACTACT AATACTTCAT CAACACTGTT TGCATGAGCT	6840
5	ACATCAAGCC CTACAACCTC GTCACCTTCA CGAAGTGTA TACCTTTCAC ACCCGTTGCT	6900
	GTACGGCCTA AAGGACGTAA TGTGATTCA GGAATCGAA TTAATGATGC ATGTGATGTA	6960
	CCAATCAAGA TATCTTCTTG ACCACTTGTT AAGCGAAGT CAATTAAGTC ATCATCTTCT	7020
10	CTGAACGAAA TCGCAATCTT ACCATTTCTA TTTATTCTTG AGAAGTTACT TAATGCTGAA	7080
	CGTTTAACGA CACCACGTTT AGTTGCAAAC ACTAAGAAGT TGTCTTCACT TTCAAGGTCT	7140
	TTAACAGCAA TCATTGTAAGT AATGACTTCA TCATTTTCAA GTTCAATAGC ATTCAGTACA	7200
15	GGAATACCTT TAGACTGTCT TGATAACTCA GGCACCTCGT AACCTTTAAG TTTGTATACA	7260
	CGACCTTTGT TAGTAAAGAA CAATACATGG TCATGTGTAC TTAAAGTTAC CAATTGACTG	7320
	ACAAAATCTT CTTCCAATGT ATTCATACCT TGAACACCAC GACCACCACG GTTTTGAGCA	7380
20	CGATATGTAG ATACCGGCAA ACGTTTAATG TAGTTATTAT GGCTTAGTGT AATTACTATT	7440
	TGTTCTTCTG GAATTAAGTC TTCGTCTCTT AAGTCTTCAA ATCCACCTAA TTGAATTTCT	7500
25	GTACGACGAT CATCACCGAA ACGATCTCTA ATTTCACTCA ATTCATCTCT AACTAACTGT	7560
	AATAACACTT CTTTCATCAGC TAAGATTGCT TCTAATTCAC TAATATAATT TAATAACTCA	7620
	TTATATTGAG CTTCAATTTT GTCTCTCTCT AAACCTGTTA GACGTCTTAA ACGCATGTCT	7680
30	AAAATAGCTT GAGCTTGTTT TTCAGAAAGT TTGAAGCGTT GTTGCAAGCT TTCCATTGCA	7740
	ACTTTATCTG TATCTGACTC ACGAATCGTT GAAATAATTT CATCGATATG GTCAAGTGCG	7800
	ATACGTAATC CTTCTAAATG GTGGGCACGA TCTTTAGCTT TACGTAAGTT GTATTGCGTA	7860
35	CGTCTTCTAA CAACTGTCTT TTGATGCTCT AAATAATGTA CCAACGCTTC TTTTAAATTA	7920
	ATAAGCTTCG GTCTACCATT TACAAGTGCA ATCATATTCA CACCAAATGA TGTGTAAGA	7980
	GGTGTGTTGTT TGTATAAGTT ATTTAAAATG AACTAGCAT TTGCATCCTT ACGCACATCA	8040
40	ATAACGACAC GCACACCAGT ACGTAAACTT GTTTCATCAC GTAAATCAGT GATACCGTCA	8100
	ATTTTCTTGT CACGAACGAG CTCTGCAATT TTTTCAATCA TACGAGCCTT ATTCAGTTGG	8160
	AAAGGAATTT CAGTGACAAC AATACGTTGA CGTCCGCCCT CACGTTCTTC AATAACTGCA	8220
45	CGAGAACGCA TTTGAATTGA ACCACGACCT GTTTCATATG CACGTCTAAT ACCACTCTTA	8280
	CCTAAAATAA GTCCAGCAGT TGGGAAATCA GGACCTTCAA TATCCTCCAT TAACTCAGCA	8340
	ATTGAAATAT CAGGGTTCTT ACTTAAGCTA AGTACACCAT TGATTAATTC TGTTAAGTTA	8400
50	TGTGGTGGAA TATTCGTTGC CATACCTACC GCGATACCTG ATGCACCATT GGCTAATAAG	8460

55

	AAATCTATTG TATCTTTATT AATATCACGT AACAGTTCAA GTGTGATTTT AGTCATACGC	8580
	GCTTCAGTAT AACGCATTGC TGCTGCGCCA TCTCCATCCA TTGAACCAA GTTACCTTGG	8640
5	CCATCAACAA GCGGATAACG ATAACTGAAA TCTTGAGCCA TACGTACCAT TGCTTCATAA	8700
	ATAGATGAGT CACCATGAGG GTGATATTTA CCCATTACGT CACCAACGAT ACGTGCTGAT	8760
	TTTTTATATG ATTTATCCGG TGTACATACCT TGTTCAATTA ATCCATATAG TATACGACGA	8820
10	TGTACTGGTT TTAAACCGTC ACGAACATCT GGCAATGCAC GAGCAACGAT AACACTCATC	8880
	GCATAATCTA AAAATGATTG ACGCATTTCA CTGGTAATAT TTCGTTTCATT TATTCTTGAT	8940
15	TGAGGTAATT CAGCCATCAA GAGTTCCTCC TTCAAAAGTT CAGTTCACAG CGCTTAGAAG	9000
	TCTAAGTTTG CATAAACTGC ATTATCTTCT ATAAATGTGC TACGGTTTTT TACAACGTCA	9060
	CCCATTAACTA TTTCAAATGT TTGGTCCGCT TCAATCGCAT CTTCAAGTTT TACTTGTAAC	9120
20	AGAGCGCGGT GCTCAGGGT CATTGTTGTT TCCCAAAATT GATCTGCATT CATTCTCTCA	9180
	AGACCTTTGT ATCGTGCAAT AGACCATTTT GGTGTTGGAT TCAATTCAGA TTAAAGTTTA	9240
	TCAAGTTCCC TATCATTGTA TACATAATAC TTTTGTTCAC CTTGTGTCAG TTTATACAAC	9300
25	GGTGGCTGTG CAATATACAC ATAGCCTGCT TCAATTAACG GTCTCATAAA TCGATAGAAG	9360
	AATGTTAATA ACAATGTTCT AATATGCGCT CCATCCACAT CGGCATCAGT CATAATGACG	9420
	ATTTTGTGAT ATCTTGCTTT CGCTAGATCA AAGTCGCCAC CGATTCTGT ACCAAATGCT	9480
30	GTGATCATTT GACGAATTTT ATTGTTATTC AAAATTCTAT CTAATCGTGC TTTTCAACA	9540
	TTTAATATCT TACCTCGTAA TGGTAAAATC GCCTGCGTTC TAGAGTCACG ACCAGATTTT	9600
	GTAGACCCCC CGGCAGAGTC CCCTTCGACT AAGAAAATCT CACATTCTTC AGGACTTTTA	9660
35	CTAGAGCAAT CGGCTAATTT ACCTGGAAGG CTTGCTACAT CTAACGCTGA TTTACGACGT	9720
	GTTAATTTCAC GCGCTTTTTT CGCAGCAACA CGTGCACGTG CCGCCATAAT ACCTTTTTCA	9780
	ACCACTGTAC GTGCGACTTG TGGATTTTCA TATAAAAATC GTTCAAAGTG CTCTGAGAAT	9840
40	AATTTATCTA CAACTTGACG CACTTCAGAA TTACCTAATT TTGTCTTCGT TTGACCTTCG	9900
	AATTGAGGAT CACCATGTTT GATAGATATA ATTGCTGTCA TACCTTCACG TGTATCTTCA	9960
	CCAGAAAGTC TATCTTTTTT TCTTTTCATA ATCTTGCTAC TTAAACCATA ACTATTTAAG	10020
45	ACACGCGTTA ATGCACGTTT GAATCCGTCT TCATGCGTAC CACCTTCATA CGTATGAATG	10080
	TTATTTGCGT AAGTTAAAAG ATTTGTGGCA TATCTGAGT TATATTGAAT CGCAATTTCT	10140
50	ACTTCAATAT CATCTTTAGA TTGATGAATA TAAATTGGCT CATCATGAAT AGGTTCTTTA	10200
	TTTTCGTTCA ATAACTCAAC GTACGATTTA ATACCGCCCT CATAGTGATA GGAGTCTTCT	10260

GCAAGCTCTC TAATACGCTG CTGTAATGTT TCATAGTTGT ATACAGTTGT CTCTGTGAAG 10380  
 ATTTCTCCAT CTGCTTTAAA ACGAAtGaCA GTACCTGTCT TAtCAGTnGT GCCAACTTCT 10440  
 5 TTTAAGTCAA ATTGAGGTAC ACCTTTTTTA TATGCTTGAT GATATATAGT CTCATTCTCTG 10500  
 TGTACATATA CTTCTAAGTC TTGTGACAAT GCGTTTACAA CTGATGAACC AACACCATGT 10560  
 AAACCACCAG ATACTTTGTA TCCGCCACCG CCAAATTTAC CACCAGCATG TAAACAGTT 10620  
 10 AAAATAACTT CGACAGCTGG ACGTCCCAT TTTTCTTGAA TATCAACTGG GATACCACGT 10680  
 CCGTTATCCG TTA CTTTAAT CCAGTTATCT TTTTCAATAA CAACTTCAAT TTGATTGCA 10740  
 TAACCAGCTA ATGCTTCATC GATACTATTA TCGACAATTT CCCACACTAA ATGGTGCAAA 10800  
 15 CCTCTCTCTG AAGTCGATCC TATATACATA CCTGGTCTTT TACGTACTGC TTCTAAACCT 10860  
 TCTAATACTT GTATTTGCCC AGCACCATAA TTATCCGTGT TGTTTACATC TGACAATGCA 10920  
 GTCACCATCG CTTTCTGTTA CTTTATAATT TCACCTTGAT TAATACGATA CAATTTAGCG 10980  
 20 TTATTCATGA TTTTCATGATC AATACCATCT ACAGATGTCG TAGTGACAAA TGTTTGTACT 11040  
 TTATGCTGAA TCGTACTTAA TAAATGCGTT TGACGCGAAT CATCTAATTC ACTGAGTACA 11100  
 25 TCGTCTAATA ATAAGATGGG ATATTCCCCA ACTTCGATAT TCATTAATTC AATTTAGCT 11160  
 AATTTAATGG ACAAAGCCGT TGTACGTTGC TGTCTTGAG AACCATATGT TTGAGCATCC 11220  
 ATGCCATTCA CATCAAACT TATATCATCT CGATGTGGTC CGAATAAGCT AATGCCTCGT 11280  
 30 TCTTTTTCTC TTTGCATATT ATCGCTAAGA ATAGACATAA TTTCTTCAAG TCGTGCCGCT 11340  
 TCATTTTGAG CATAATCAAA TTTAAGACTA GGTAAATAAT TCAGCGACAA CGCTTCTTTA 11400  
 TCATTGTGA TACCAGCATG AATCGGTTTA GCTAACGACT CTAGCTCTTG AATAAAATGT 11460  
 35 GCACGTTTAT CAGTTACTTT CATTGCATAT TCAGCAAAC GCTGATTTAA TACTTCCAAC 11520  
 ATTGTTAAGT CCTTTTTTTG GCCTAATTGT AACTGCTTTA AGTAATTATT CTTTGTCTTT 11580  
 AAAATACGTT GGTATTGAGC TAAATCATTT AAGTAAACAG CAGAAATTTG GCCCAACTCC 11640  
 40 ATATCTATAA AGCGTCGTCT TATTtGrGGr GAGCCTTTTA CAATATTCAA ATCTTCTGGC 11700  
 GCAAATAGAA CCACATTGAG GTGTCCAATA TATTGAGTTA GACGACTTTG CTCTAAGTn 11760  
 ATTCACTTTG GACTTGTTTA CCTTTnTTAG TTATAAACAT TGTTAATGGG CATCGTGCCG 11820  
 45 TGT 11823

(2) INFORMATION FOR SEQ ID NO: 137:

- 50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 692 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

5 ATAATTATTA ACATGGTGTG TTTAGAAGTT ATCCACGGCT GTTATTTTGG TGTATAACTT 60  
 AAAAATTTAA GAAAGATGGA GTAAATTTAT GTCGGAAGAA GAAATTGGG AAAAAGTGCT 120  
 TGAAATTGCT CAAGAAAAAT TATCAGCTGT AAGTTACTCA ACTTTCCTAA AAGATACTGA 180  
 10 GCTTTACACG ATTAAAGATG GTGAAGCTAT CGTATTATCG AGTATTCCTT TTAATGCAAA 240  
 TTGGTTAAAT CAACAATATG CTGAAATTAT CCAAGCAATC TTATTTGATG TTGTAGGCTA 300  
 TGAAGTTAAA CCTCACTTTA TTAATACTGA AGAATTAGCA AATTATAGTA ATAATGAAAC 360  
 15 TGCTACTCCA AAAGAAACAA CAAAACCTTC TACTGAAACA ACTGAGGATA ATCATGTGCT 420  
 TGGTAGAGAG CAATTCAATG CCCATAACAC ATTTGACACT TTTGTAATCG GACCCGGTAA 480  
 CCGCTTTCCA CATGCAGCGA GTTTAGCTGT GGCCGAAGCA CCAGCCAAAG CGTACAATCC 540  
 20 mTTATTTATC TATGGAGGTG TTGGTTTAGG aAAACCCAT TTAATGCATG CCATTGGTCA 600  
 TCATGTTTTA GATAATAATC CAGATGCCAA AGTGATTTAC ACATCAAGTG AAAAATTCAC 660  
 AAATGAATTT ATTAAATCAA TTCGTGATAA nA 692

25 (2) INFORMATION FOR SEQ ID NO: 138:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7900 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

35 ATACTGTAGC GCAAATTTCA CAATGGCATG TTATAGAAGA TTTAGTTACG AATGAATTAG 60  
 GTATTAGTAT TTTACCAACA TCAATTTTCAg CAAGCTAA TGGAGATGTG AAGCTGTACG 120  
 40 CATTGAAGAT GTCATGTAC ATTGGGAATT AGGTGTTGTT TGGGAAGAAGG ATAAACAATT 180  
 AAGTCATGCC ACAACGAAAT GGATAGAATT TTTGAAAGAC CGTTTAGGCT AACATATTAA 240  
 TAAAGCACTC ATTATTTAAG GCGCATCATT ACGTGGGTCA TTGAAATAAT GAGTGTTTTT 300  
 45 TTGTGAAAT GAAGTGAAAT TTAGAGAGCG TTTCATAGA AAATAGTAAT ACAAACTATA 360  
 AAAAAAGAGT ATTTTATAT TGTGTACGCC ATCTTTATA TAGTTATTGT AACAAATTTAG 420  
 ACATATTTAG AAAGGGATGG CGCCATGCAC AAAGTCCAAT TAATAATCAA ACTACTACTA 480  
 50 CAACTAGGAA TCATCATTGT GATTACTTAT ATTGGCACAG AAATTCAAAA GATTTTTCAT 540

55



EP 0 786 519 A2

	ATTGTACCGC TAACTTGGGT AGAAGACGGT GCAAACCTTT TATTAAAGAC GATGGTCTTT	660
	TTCTTCATAC CGTCAGTTGT AGGtATTATG GaTgtgCTTC CGAAATTACG CTAAATTATA	720
5	TACTCTTTTT CGCAGTCATT ATCATAGGAA CATGTATCGT TGCATTATCT TCAGGTTATA	780
	TTGCTGAAAA AATGTCyGtT AAACwTAAAC ATCGTAAAGG TGTAGACGct TATGAATGAT	840
	TACGTGCAAG CCTTATTAAT GATTTTGTTG ACTGTCTGTTT TATATTATTT CGCTAAAAGG	900
10	TTACAACAAA AATATCCGAA CCCATTTTTG AATCCAGCAT TAATTGCATC TTTAGGAATT	960
	ATTTTTGTCT TACTTATCTT TGGAATTAGT TATAACGGGT ATATGAAAGG TGGCAGTTGG	1020
	ATCAACCATA TTTTAAACGC AACGGTCGTA TGTTTAGCGT ACCCACTTTA TAAAAATAGA	1080
15	GAGAAAATTA AAGACAATGT CTCTATCATT TTTGCAAGTG TATTAActTGG CGTCATGCTG	1140
	AATTTTCATGT TAGTGTTCTT AACACTTAAA GCATTTGGCT ATTCTAAAGA CGTCATTGTA	1200
	ACGTTATTGC CCCGATCTAT AACAGCCGCA GTAGGTATCG AAGTGTCA CA TGAActAGGT	1260
20	GGTACAGATA CGATGACCGT ACTTTTTATT ATCACAACGG GTTTAATCGG TAGTATTTTA	1320
	GGTTCGATGT TATTAAGATT TGGAAGATTT GAATCTTCTA TCGCCAAAGG ATTAACGTAT	1380
	GGGAATGCGT CACATGCATT TGGCACAGCT AAAGCACTAG AAATGGATAT TGAATCCGGT	1440
25	GCATTTAGTT CAATTGGGAT GATTTTAACT GCAGTTATTA GTTCAGTGTT AATACCTGTT	1500
	CTAATTTTAT TATTCTATTA ATTTAGATAT TTAAAATGAT AGACAGAAAG GGAGGCTATT	1560
30	AGTAATAATG GCAAAAATAA AAGCAAATGA AGCATTAGTT AAAGCATTAC AAGCaTGGGA	1620
	TATAGATCAC TTGTATGGTA TTCCAGGAGA CTCAATCGAC GCATAGTCGA TagTTTACGT	1680
	ACAGTGAGAG ATCAATTTAA ATTTTATCAT GTACGTCATG AAGAAGTAGC AAGCTTAGCG	1740
35	GCTGCTGGTT ACACAAAATT AACTGGTAAA ATCGGTGTGG CATTAAGTAT CGGTGGCCCT	1800
	GGTTTAATTC ATTTATTAAA TGGTATGTAT GATGCCAAAA TGGATAATGT ACCGCAATTA	1860
	ATATTATCTG GACAAACGAA TAGTACAGCA CTTGGAACGA AAGCATTCCA AGAAACAAAT	1920
40	TTACAAAAAT TATGTGAAGA TGTAGCCGTT TATAATCACC AAATTGAAAA AGGTGACAAT	1980
	GTGTTTGAAA TCGTTAACGA AGCAATTCGT ACGGCATATG AACAAAAAGG TG TAGCTGTT	2040
	GTTATTTGTC CTAACGACTT ATTAActGAA AAAATTAAAG ATACAACGAA TAAACCAGTA	2100
45	GATACATCAA GACCAACAGT AGTATCACCA AAATATAAAG ACATCAAAAA AGCGGTTAAA	2160
	CTAATTAATA AAAGTAAAAA GCCTGTCATG TTAATTGGTG TAGGTGCGAA ACATGCGAAA	2220
	GATGAGCTAC GTGAATTTAT TGAAATGGCT AAAATTCCTG TCATTCATT CATTACCAGCT	2280
50	AAAACAATCT TGCCGGATGA TCATCCATAT AGTATCGGtA ACTTAGGTAA AATCGGTACC	2340

55

	CCATATGTGG ATTACTTACC TAAGAAAAAT ATTAAAGCCA TTCAAATTGA CACAAATCCT	2460
	AAAAATATCG GACATCGTTT CAATATTAAT GTAGGAATTG TTGGAGATAG TAAAATTGCG	2520
5	TTGCATCAGT TAACTGAAAA TATTAAACAT GTTGCTGAAA GACCATTCTT AAACAAAACG	2580
	TTAGAACGTA AAGCGGTTTG GGATAAATGG ATGGAACAAG ATAAAAATAA TAATAGTAAA	2640
	CCATTACGTC CAGAACGATT AATGGCATCA ATCAATAAAT TTATTAAAGA TGATGCAGTG	2700
10	ATTTGAGCAG ATGTAGGTAC AGCAACAGTT TGGTCAACTC GATACTTAAA CCTTGGTGTA	2760
	AATAACAAGT TCATCATTTT AAGTTGGTTA GGTACAATGG GTTGCGGTCT TCCAGGTGCA	2820
	ATTGCATCAA AAATTGCATA TCCAAATAGA CAAGCCATCG CAATTGCTGG TGACGGTGCA	2880
15	TTCCAAATGG TAATGCAAGA CTTGCTACA GCAGTACAAT ATGATTTACC TTAACTGTA	2940
	TTTGTACTTA ATAACAAACA GTTAGCATTT ATTAAATATG AACAAACAAGC AGCTGGTGAA	3000
	TTAGAATATG CAGTTGATTT TTCTGATATG GATCATGCAA AATTTGCTGA GGCAGCAGGT	3060
20	GGTAAAGGTT ATACAATTAA GAGTGCTAGC GAAGTAGATG CTATAGTCGA AGAGGCATTA	3120
	GCACAAGATG TACCAACGAT TGTAGATGTA TATGTTGATC CTAATGCTGC GCCATTACCA	3180
	GGTAAAATTG TAAATGAAGA AGCGCTGGT TATGGTAAGT GGGCATTTAG ATCAATTACT	3240
25	GAAGATAAAC ATTTAGATTT AGATCAAATT CCACCAATTT CAGTGGCAGC AAAACGTTTC	3300
	TTATAACTGA TTAAAGGTT ATCACAATTG AATTGAACTA TAAAAACGGT AATTTCTATT	3360
	TCAACAAAT GGAATTGCC GTTTTGTTTA TTTATCACAA ATGATCGTAC TGAATTGATG	3420
30	ATAAAATTGT GAAAAAGTTG TTGAAAACGC TTTTACAAAT ATGTATAATA GCTATGAATT	3480
	AGATATCACT TCGTGTTAC TGGTAATGCA GGCATGAGCA AACAAACGCA CTATGAGAAT	3540
35	AGTCTTGTTT GTTCATGCCT GCTTTTTTTG TACATGGAAG CGGAAATTGA GATAGGGGAT	3600
	GTTTETATGT TTAAGAAATT GTTTGGACAA TTGCAACGTA TCGGTAAAGC ATTAATGTTA	3660
	CCTGTTGCGA TTTTACCAGC AGCTGGTATT TTATTAGCGT TTGGTAAACG AATGCACAAC	3720
40	GAACAATTAG TAGAAATTGC ACCATGGTTA AAAAACGATA TCATTGTAAT GATTTGTCG	3780
	GTCATGGAAG CAGCAGGACA AGTTGTATTT GATAACTTGC CATTATTATT TGCAGTTGGT	3840
	ACAGCACTTG GATTAGCAGG AGGAGACGGT GTTGACGAT TAGCAGCGCT AGTAGGTTAC	3900
45	TTAATTATGA ATGCAACAAT GGGGAAAGTG TTGCACATTA CAATTGATGA CATTTTCTCA	3960
	TATGCCAAAG GGGCAAAGA ATTAAGTCAA GCAGCGAAAG AACCAGCACA TGCTTTAGTA	4020
	TTAGGTATTC CAACGTTACA AACGGGTGTG TTTGGTGGTA TTATCATGGG TGCTTTAGCC	4080
50	GCATGGTGTT ACAACAAATT TTATAATATT AACTACCAC CATTTTTAGG ATTCTTTGCA	4140

55

EP 0 786 519 A2

	AGCTTTGCGT GGCCACCAAT TCAAGATGGA TTAAATAGTT TATCGAATTT CTTATTAAAT	4260
	AAAAATTAA CATTAAACAAC GTTTATATTC GGTATTATTG AACGCTCATT AATTCCATTT	4320
5	GGTTTACATC ATATTTTCTA TTCACCGTTC TGGTTTGAAT TCGGAAGTTA TACAAATCAC	4380
	GCAGGTGAAT TGGTTCGTGG TGACCAACGT ATTTGGATGG CACAATTGAA AGATGGCGTA	4440
	CCATTTACTG CTGGTGCATT TACTACTGGT AAATATCCAT TTATGATGTT TGGTTTACCA	4500
10	GCGGCGGCAT TTGCTATTTA TAAAAATGCA CGACCAGAAC GTAAAAAAGT CGTGGGTGGT	4560
	TTAATGTTAT CAGCAGGATT AACTGCATTT TTAAGTGGTA TCACTGAGCC ATTAGAATTT	4620
	TCATTCTTAT TTGTAGCACC AGTACTTTAT GGAATTCACG TATTATTAGC TGGTACATCA	4680
15	TTCTTAGTAA TGCATTTATT AGGCGTTAAA ATTGGTATGA CATTCTCAGG TGGTTTCATA	4740
	GATTATATTT TATATGGTTT ATTAACTGG GATCGTTCAC ACGCATTATT AGTTATTCCA	4800
	GTCGGTATTG TATATGCTAT CGTGTATTAC TTCCTATTCTG ACTTTGCAAT TCGTAAGTTT	4860
20	AAATTGAAAA CACCAGGTCG TGAAGATGAA GAAACTGAAA TTCGTAAGTC TAGTGTCGCA	4920
	AAATTACCAT TTGATGTCTT AGATGCAATG GGTGGAAAAG AAAACATTAA ACATTTAGAT	4980
25	GCATGTATTA CACGTCTACG CGTAGAAGTG GTTGATAAAT CAAAAGTAGA TGTAGCAGGT	5040
	ATTAAAGCTT TAGGCGCATC AGGTGTATTA GAAGTTGGAA ACAATATGCA AGCTATCTTT	5100
	GGTCCAAAAT CAGATCAAAT TAAACATGAT ATGGCCAAGA TTATGAGTGG TGAAATTACG	5160
30	AAACCAAGTG AAACGACAGT GACTGAAGAA ATGTCAGATG AACCAGTTCA CGTAGAAGCA	5220
	CTTGGAACAA CAGACATCTA TGCACCAGGT ATCGGTCAAA TCATTCCATT ATCAGAAGTA	5280
	CCTGATCAAG TATTCGCTGG TAAAATGATG GGTGATGGTG TTGGCTTTAT CCCTGAAAAA	5340
35	GGTGAAATTG TAGCACCGTT TGATGGTACA GTGAAAACAA TCTTCCCTAC GAAACATGCG	5400
	ATAGGATTAG AATCTGAAAG TGGCGTCGAA GTACTTATTC ATATTGGTAT CGATACAGTG	5460
	AAACTGAATG GTGAAGGATT CGAAAGTCTG ATTAACGTTG ATGAAAAAGT AACACAAGGT	5520
40	CAACCATTA TGAAGTGAA TTTAGCATAC TTGAAAGCAC ACGCACCAAG CATCGTTACA	5580
	CCAATGATTA TTACAAATCT TGAAAATAAA GAAGTTGTCA TTGAAGATGT ACAAGATGCT	5640
	GATCCAGGTA AGCTAATTAT GACAGTCAAA TAATGATTAA AAATGAAACA GCATATCAAA	5700
45	TGAATGAAGT TTTAGTCATT CGTAGTGGT ATGCGAAGTA GCGAGTTGAA AGAGAATACG	5760
	TTACAAAAGG CAGTAGCTTA AAATGAAGCT ACTGCCTTTT TAGTGCGCAA TGATGTATAG	5820
	CAGGTGTGTT GATGTATAATA AGTTAAATAT TAGTGTTAGA TATAGAAAAC ATTGCTTATG	5880
50	TTTTTGTCAC ATTTTAGAAA AATGCATCTT CGCGACTAGC CAAATTAATA GTCTCATTGA	5940

55

	AATAAAATTAA CATGATTTTA AATCTATTTG TAAGATAAGG AGATTTGTCA TTATGACAAC	6060
	AGAAGGTCTA TTAGTTGCAG AGAAAGAAAT CGAAGTGAAT GGTACGACA TTGATGCGAT	6120
5	GGGTGTCGTT AGTAATATCG TTTATATTAG ATGGTTCGAA GATTTGAGAA CAGCGTTTAT	6180
	TAATCAGCAC ATGAATTACT CAACAATGAT CAATCAAGGC ATTTACCTA TACTTATGAA	6240
	AACGGAAGCA GAGTATAAAG TACCTGTCAC AATACATGAC AAACCAGTAG GTCGTATTTA	6300
10	CTTAGTTAAA GCAAGCAAGA TGAAATGGGT GTTTCAGTTT GAAATTGTGT CCGCACATGG	6360
	CGTGCATTGT ATTGGTACAC AGACAGGCGG TTTTACAGA TTGAGTGATA AGAAGATAAC	6420
	CTCTGTGCCA CAAGTGTTTC AAGACATTTT AGCAACAAAA TAATGACTTC ATTTTAAAAAT	6480
15	ATAAAAAGTA AGAAGGTGTT CGAAATGGTT AAGCAATTAA ATAGTGTCGA AGCATTCCGT	6540
	GAATTTATTC ATCAATATCC GTTAGCAGTT GTACATGTCA TGCGCGATCA GTGTAGCGTG	6600
	TGTCATGCCG TTTTACCACA AATTGAAGAC TTGATGCAAT CATATCCCAA TGTGCCATTA	6660
20	GCTGTGATTA ATCAAAGTCA GGTGGAAGCT ATTGCTGGAG AATTAAATAT TTTCaCTGTA	6720
	CCTGTGGATT TAATTTTAT GAATGGAAAA GAAATGCATC GTCAAGGGCG TTTTATCGAT	6780
	ATGCAACGTT TTGAACATCA TCTTAAGCAA ATGAATGATA GTGTAAATAA CGATGTCGAT	6840
25	GAGCATTAAT ATCGCAAATG ATTAGCATTG CTAAGATTAT GTAGACATCA TAACTTATTT	6900
	CCCAGTAAAT ATTGGTAGTA ATTAGAATCA GCATGGTACA GTAGAACTAT AGTAGAAATC	6960
	ATCAAAGAGG AGTGACGACA AATGCGTAAA AAATGGTCTA CACTTGCGTT TGGATTTTAA	7020
30	GTTGCAGCAT ACGCACATAT TAGAATTAAA GAAAAACGCA GTGTGAAAAG TTATATGTTA	7080
	GAACAAGGTA TACGATTATC TAGAGCTAAG CGTCGTTTAA TGTATAAAGA AGAAGCGATG	7140
35	AAAGCATTAG AAAAAATGGC GCCACAGACA GCAGGCGAAT ATGAGGGAAC CAATTATCAG	7200
	TTTAAGATGC CAGTAAAAGT GGATAAGCAC TTCGGTTCAA CCGTTTATAC CGTTAACGAT	7260
	AAACAAGATA AGCATCAACG CGTTGTATTA TATGCACATG GAGGCGCATG GTTCCAAGAC	7320
40	CCACTCAAAA TTCATTTTGA ATTTATTGAT GAACCTGCAG AAACACTCAA TGCTAAAGTC	7380
	ATCATGCCAG TATATCCGAA GATTCCGCAT CAAGATTATC AAGCGACGTA TGTGCTTTTT	7440
	GAAAAGTTGT ACCATGATTT ATTGAATCAA GTAGCAGATT CTAAACAAAT CGTTGTAATG	7500
45	GGTGACTCTG CGGGCGGTCA AATTGCTTTA TCATTTGCTC AATTGTTAAA AGAAAAACAT	7560
	ATTGTGCAAC CAGGACATAT TGTATTAATT TCACCAGTTT TAGATGCAAC GATGCAGCAT	7620
	CCTGAAATTC CTGACTACTT AAAGAAAGAC CCAATGGTAG GTGTGGATGG CaGTGTGTTT	7680
50	TTAGCTGAAC AATGGGCAGG GGACACACCT TTAGATAACT ACAAAGTATC ACCAATTAAT	7740

55

CCAGATGCTT TGAACCTATC GCAATTGTTG AGTGCGAAAG GTATCGAACA TGACTTTATA 7860  
CCTGGATATT ACCAATTCCA TATTTATCCA GTATTTCCGA 7900

5 (2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1984 base pairs  
(B) TYPE: nucleic acid  
10 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

15 GTCTAAATAA ACAAATTAT CATTGATTaC TGAACCTGGCA TTTCGAAGTA ATGCTTCAAT 60  
ATCATTGCAA TATTTCTTCA ATTTATGATT GTGAAATAAT TCTTGCATCA AAAATGGTCT 120  
TTGGTCACAT GAATGTGCAT CTGAAGCTAC AAAATGAGCC AAATTACATT CTATAAATTG 180  
20 TAATGATAAC TTTTGAATGT TTTTACCAA TCCACCAACT AAAGAACTCG ATGTTAATTG 240  
ACTCAGTGCC CCATTTGCAA CCAATTCATA TAATATTTCC GGATTTTGG CGATACTTCT 300  
ATTTCTTTCA GGATGTGCAA TGATTGGTAT GTAACCTCTC GATTGTATTT CAAAAACAA 360  
25 TTGTTTTGTA TAATGTGGTA CTTGCCCCGT TGGAAATTCA ATTAATAAAT ATTTGGAACG 420  
ATTAATACCT TGAATACTAC CATTATCTAA GCCTTTCAGA ATCGAATCTG TAATTCTAAT 480  
TTCTTGCCCG GGAAATAATT TAATATCCAA TGCTTGAAC TCTGGATGCG TTCTTAATC 540  
30 CGCCAATTTC ACAAGCACTT GTTGAAATGT ATTATCATAT CTCGGATGCA AATGATGAGG 600  
TGTCGCTACA AACTTGTGA CACCTTCATC CTTAGCTTGC TTTAATAGTG CAATACTCTT 660  
35 TTCAATTGTT TTAGGACCAT CATCTATATC AACTAATATA TGGTTATGAA TATCAATCAT 720  
GATTCATCAG TCCATAATA TGCATAGTAA CTAGCACTTT TATCTTTAGG CATTCTATTT 780  
AAGACTACAC CTAATAATTT AGCACCTGTT GCTTCAATAA GTTCTTTTCC TTTTAACT 840  
40 TCATCTCTAT TATTATTTTC CGAATTAAC ACGTAGACAA CATTGCCGGT AAACCTTGAA 900  
AATAATTGCG CATCTGTAAC TGTGTTCACT GGTGGCGTAT CGATAATTAC AAAGTTATAA 960  
TTCATCAATA ATGTGTCATA CAAATTTGCA AATGCCCTTG ATGTAATTAA CTCTGACGGA 1020  
45 TTCCGTGGGA TTGGCCGAGA CGTCAAGACG TCTAAATCTT GAATTTCACT TGAGATAATA 1080  
CTGTCTTGAT AAGTTGACCA ATTTAGCAAT AAACCTGATA GGCCTTCATT GTTTGGCAAA 1140  
TTAAAAATAT AATGCTGCGT AGGTTTACGC ATATCCCCGT CTACGATTAG TGTTTTATAA 1200  
50 CCTGCTTGCG CATATGCAAC TGCTAAATTT GCTGCAATTG TAGACTTACC TCGCCTGGT 1260

55

GATCTTATGC CTCGAAATTT CTCGCTAATA GGTGACTTTG GTTGTTTCATG GACAATTAAA 1380  
 CTTGATGTAC TTCyTCGTGT ATTTCGTCATG GTAATTCCTC GTAAATTAAA ATTTTTGTAT 1440  
 5 TGAACCTAAA ATAGGTAATC CTAGTTGCGA TTCAACATCT TCTTCTGTCT TAATACGCTT 1500  
 ATCTAATAAT TCTTTTAAGA AAATAATCAA TATTGCTAAA ACAATACCAA CAATAATGCT 1560  
 GATAACTAAG TTGACAGATA CTATTGGAGA TACTTTTACA GCATTATCAT GTGCTGAGGA 1620  
 10 AAGTATCGTA ACATTATCAA CACTCATAAT TTTAGGCATG TCATGAGCAA AAACCTTTAGA 1680  
 TATTTTATTA ACAATTTTGT CAGATTGAGA TTTATTCCCA GTGGTAACTG ATACAGTAAT 1740  
 AATTTGAGAG TTTGTTTGAT TGGTTACTTT TAAAAATGAA TTCAACTCAG CTGTTGAATA 1800  
 15 CTGACCATCA AnTTCTCTAG ATACTTTATC TAGAATTCTA GGACTTTTGA TAATTTCCGT 1860  
 ATATGTATTA ACAGACTGCA AACTACTTTG AACATTTTGG AAAGCTAAAT CACTTGAGGA 1920  
 CTTTTTCATG TTCACTAATA TTTGAGTAGA AGCAGTATAT TTGTCAGGCA TAACAAAAAA 1980  
 20 GGTT 1984

(2) INFORMATION FOR SEQ ID NO: 140:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6272 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

CAAATCCCTT GGTGATGAtA AAtGtATTGC TGTGTAGCCA AATAATCTTC GTATATATGA 60  
 35 CTGACGTTCA ACAACAGCTT GCAATCGTTT CGTTGGTACA GTTACTTTCT TCTTGTTAAA 120  
 GAGACCATAT TCAATTTTAA GTTGCTCATT TTCAAGCATC ACCGAAAAGC CATAAAATCT 180  
 TATCATTGTT ATAATCGTTC CAATAATATA TGCCACTATT AATACTAGTA AAATGATGAT 240  
 40 TAATACTGAA ATACTTACAA TTTGAACCCA TTGACTAATT TCATGATTTA GCTTCGACCA 300  
 TGGGATCAAC TCTCTTACAG CCCCCTAAAT CGGTAATAAA GCTGCTAACG TTACACCAAT 360  
 GGCGCCACTG GTCATTGCCA TAAATAGTGA TTCTTTAAAA TTCATCTGAT ATATAGGAAT 420  
 45 GCGTTTATTT TTCTGATTAA GCATACTATC AGTGTCTGTC ACTTCATCTA AGCGACCTTC 480  
 TGCGATGTCT TCCACATTAC CTTCAATGTC ATGATTACAG TTGTCATTCT TCTCAGCACT 540  
 AGACTTTTGC GCCACTTCTG TCTTCAACTC TGTTTGCAAT TGATCAATAT ATCGTTCAAG 600  
 50 ATATTCACCT TGTTTTTTTCG AAATAACACT TAAGACAATA CCATCACTTG GTGTTTTGAT 660

55

	AATACGTTTT ATATTTAATT CTTTACGCTT TTTATTAAAA ATACCTGTTG TTAAAAATGAA	780
	ATAATTATCC TCAATCCAAT ATCGCGTGTT CATAATTCGG ACAATTTGAG AAATGTATGA	840
5	TATTAAAAAG AATACAAATA CAATACCTAT CCATAAATAT GATTCGGGAT TCGTATAATC	900
	AAAATCTTTC AATTGAAAGA TAATGAAAAT AAAAAAGACG ACTATGTTTT GTTTGATAGC	960
	ATTGATTATG CCATTAAAAT ATGAAATCGG ATGTAATTTT TGAGGTTTCAG ACATCACTTT	1020
10	CAACCCCTCT CAAATTCGAC ATAGTTCTCT CTTGATTAT TTAAACATCG TCATGAGACA	1080
	TCATCGGTAA ATAAATAGTA TGACCTGCAG TCATAAATCC AACTTTATAC AAATTAAGCA	1140
15	CTTTACTAAT TGGATTAGAT TTAATCGACA AGTATTGTAA ACGTTCAATT CGACTCGTTT	1200
	CTTCTTTATA TATAAAAAAT GATGTACGAT ATTGTACACT TAGTTGATCA ACTTTATAAA	1260
	AGCGACAATG ATATTGCCAT AAAGGCTTAA TAAATAATTT TAATGTACTC AGAGCACCTA	1320
20	AAACCAACAA AATATAAAGT AAGTAATGTG GCCATTCAAA TCTTAACCAT ATAAATAAAA	1380
	AAATGACATA CACAGCTACA CTCAATATAA ATTCTAAGCC ATTCGTAATG TAGTAATACA	1440
	ACAATGCTGA CTTAGGACTC TTAGTCAACT TAGTATAATC TGACATATAC CCCTCTCCCC	1500
25	AAATAAAAAA TTATACGGAT TTATAATCTA TTTCATTTTA TTTTATATG ATGATAATTA	1560
	TAGCATATGG AATATTTTCAT GCTAATTTAT TCTTCCTAAA GGTACATCTA AAAATTTAAT	1620
30	TAAGCAGAAA GTGCTTGAAT TGCTAAAAAG ACACCATGTT ATAATTTTAT CAACATGATG	1680
	CCTTTCATCT ATAATCAATC TTTCATCTTA TCAAGAGCGA TATTTAGTTC AAGCACATTC	1740
	ACATAATCAT TTGTTAACAC ACCACGCTGC TTACGATGTT GAATCAAGTC GGCCACTCTT	1800
35	GAAGTAGATA CATGACGAGC ATCAGCAATA CGAGGTGCTT GCTTCAATGC ATTTTCGACC	1860
	GTAATATGCG GATCTAAGCC CGACCCAGAA CTTGTTGCAG CATCTATTGT TACATTTGAA	1920
	TTCCCAAATT TAACATGATG TTTCATGCGT GCTATTAATT CGGTGTTTCC ATTCGATTCA	1980
40	TTACTTCCAC CTGAAGATAC GCCGTTTTTA TATAATTTTT CAGGATTCAT ATTATAATCA	2040
	ACTGCACTCG GTCTCCCGTG AAAATATCGT GTCTCTGTCC AGTGCTGTCC AATCAATTTT	2100
	GATCCAAC TAACGATTGTC ATACGTAATT AACTGCCAT TTGCTTGTG ATAAAAAAT	2160
45	ATTTGACCAA TTAACGTGAT AGCTAACGGG AATAAAAATC CACATAATAC CATAGTTATT	2220
	ATCGTTAAAC AAATACTATT TCTTATCGTA TTCATGGTAC AGGCTCCTTC CTCTTTACAC	2280
50	AAAAAATTGT ACAATCATAT CTATTAAATTT AATGCCTAAA AACGGGACGA TTAATCCACC	2340
	TAATCCATAA ATCAACATAT TATTTATAAA GATTCTATCA ATGCTGTAAC CCTTTACTTT	2400
55	TACACCTTTC ATGGCAATTG GAATTAAGGC AACAATGATT AATGCATTGA ATATCAAAGC	2460

	AATTGTTGAC ATCATTAGTG CAGGTAAAAT TGCAAAGTAT TTTGCTACGT CATTAGCCAA	2580
	ACTAAATGTC GTTAATGCAC CTCTCGTCAT TAATAATTGT TTGCCTATTT TTACAACCTC	2640
5	TATTAACTTT GTAGGATTCG AATCTAAATC AATTAGATTA GCTGCCTCTT TAGCACTAAT	2700
	TGTCCCTGAG TTCATAGCTA ATCCTATATT CGCTTtGTGc tAGCGCAGGT GCATCATTTG	2760
10	TACCATCTCC TGTATCGCA ACAATATGGC CTTTCGCTTG TTCATCTTTG ATGACTTTAA	2820
	TTTTATCTTC GGGTTTACAC TCTGCAACAA ATCTATCAAC CCCGGCTTCT TTTGCAATTG	2880
	TAGCTGCTGT TAAAGCATTa TCACCTGTAC ACATAACTGT TTCAATCCCC ATTTTTCTCA	2940
15	ATTcAGTAAA TCGTTCTACA AGACCATCTT TAATCACATC TTTTAAATAA ATCACGCCAA	3000
	GCATGACATT GTTTTCAATG ACTATTAAATG GnGTGCCACC TTTACTCGAT ACATCCATAC	3060
	AGAGAGACTC AATATTAAAGa GGAATATTGC CTTGTTGTG TTTGACAAGA TTTATCATAC	3120
20	TATTAGGTGC ACCTTTGAAT ACCGATATTT CATTTGTAAT GATTCCGCTC ATTCTAGTTT	3180
	CAGCTGTAAA AGGCTTATAT GTGCCATCAA TGTCTTTAGG CAGCTCATTT ATATACATcT	3240
	GcttCGCTAA TCGTACAATA CTTTTTCCTT CTGGCGTATC ATCGTAGATT GATGACATAT	3300
25	AAGCAGCGAC TATCAATTTT TCAAGCATTT GTTGATTcAC TGGTAAAAAT TCACTAGCGA	3360
	TTcGATTGCC ATAAGTGATT GTGCCTGTCT TGTCTAAAAT CATTACATCG ACATCTCCAC	3420
30	ATACTTCTAC AGCACGCCCA CTTTTCGCTA ATACATTGAA TTGAGTAACA CGATCCATGC	3480
	CTGCAATACC AATCGCCGAT AACAAACCAC CGATTGTCGT TGGTATTAAA CATACTGTTA	3540
	ACGCAATGAG CATCGCAATA GGTAAAATTA AATGCAGGTA AGATGCTATT GGATATAACG	3600
35	TTACAATAAC GACTAAAAAT ATAATTGTTA ACGTTGTTAA TAATGTAAAA AGTGCAATTT	3660
	CATTTGGTGT TTTATTTCTT TCCGCCCTT CAACTAAGGC AATCATTTTA TCTAAAAAAG	3720
	ATGTAcnCGC TTCACTCTCA ACACGTATTT CTAACCAATC AGATGTTACA AGTGTAACGC	3780
40	CAATGACTCC ATCAAAATCG CCACCTGATT CTTTTATCAC AGGTGCAGAC TCACCAGTAA	3840
	TTGCAGATTC ATCAACGGTT GCTAATCCAT TTATTACAAC GCCATCAGCA GGGATTGTTT	3900
	CTCCATTTTC TACCCGAATA TTTTGTCCGG CTTTAACTC TGTGGCGTTC ACTATCCGAT	3960
45	ACGCACCATT TTCTTCTATC AATCGAGCAG TTAAATTTGA TTGTGCTTGT CTAAACTAT	4020
	CAGCTTGCGC TTTTCCACGA CCTTCAGCAA AGGCTTCTGA AAAATTAGCA AACAATATAG	4080
50	TTATTAATAA TATGATAAAA ATTGTAATCA AATAACCTCG CGATAGATAG CTAGTTCCAA	4140
	ATATGTCAGG AAAACATATT AATATCAACG TTAAATCAT TCCAACCTCA ACGACAAACA	4200
	TTATCGGATT TTTTATTAAT TGTTTAAGAT TCAGCTTATA AAAACTCATT TTCAAAGCTT	4260
55		



	TTTATTTTAA AGTTAAAAAT TCACCAATAG GACCAAGTAA TAGTACTGGA ATAAATGTCA	4380
	AACCACTTAG TAAAACGATA AATACGATTA GTGATACGCC AAAATAAGGT TTATCAATCG	4440
5	CTATTGTATA TTTATCTTGA TGGTATGATT TTTTATTAC TAAACTTGAT GCAATCATT	4500
	ATTGCAAAAT AATTGGTATA TAACGAGAAA GCAACATAAT GATTCTGTGA GAGATATTCC	4560
10	AGAATGTTGT ATCATCTTTC AGTCCTTCAA ACCCTGATCC ATTGTTGCGA GCAGCTGATG	4620
	TCATTTCATA CATAACTTGT GAAATACCAT GAAAAGACGG ATTCGTTATa CTTtCACTTG	4680
	CTCCAGGAAT CATAAAAGCA AGTGCTGAAA ATACTAAAAT TAAAATTGGG TGTATGAGAA	4740
15	AGACTAAGAC AATACATTTT ATTTCACGGG CGCCAATTGG CATATTTAAA TATTCTGGTG	4800
	TTTTACCAAC CATCAAACCTG CATATAACA CCGTCAGTAA GACAAATATC AATAAATTCA	4860
	TGAGTCCTAC GCCTTCGCCA CCAAATACAA CATTTAGCAT CATTAAATACC ATTGGTCCTA	4920
20	ATCCACCTAT AGGCGTTAAG CTATCATGCA TGTATTAAAC AGAACCCGTT GTAAATGCCG	4980
	TCGTAATAAC TGTAATAGT GCTGACAAAC CTGCTCCAAA CCGTACCTCT TTACCTTCCA	5040
	TATTCGGTCC ATAAATGCCT AAATTCGCTA GTATTGGATT ACCACGATAC TCACTCCACA	5100
25	TAGTTAATGT AAGAATTGCT ATAAAAATGA AAAACATTGC GACAAATAAT ATCAACGCAT	5160
	GACGATGTAC TCGTTTACCA TGTCTACTTA ACATGCGACC AAATAAGAAC AACATTGACA	5220
30	TAGGAAGTAA CATCATACTG CCCATTTCTA TAAAATTGCT CCAAATATTT GGATTTTCAA	5280
	AAGGTGTTGC AGAATTTTCT GCTAAAAATC CTCCACCATT CGTACCAAGA TGTTTTATTG	5340
	ATTCAAGTGA TGCAATAGGT CCAAATGCAA TATGTTGAAT ATGTCCGCTT AAAGTCCGAA	5400
35	TCATTAAATT AGCATGCAAC GTTGTGGTA CaCCTTGAGT CATCAATAAA ATACTAATTA	5460
	AACATGATAA TGGTAAAAGT ACTCGGACAA TAAACCGAAC AATATCTTGA TAAAAATTAC	5520
	CAATGATATT AGTTAATCCA GTTAAACGTC TCAACATCGC TATACAAACG GCGTAACCTG	5580
40	ATGCACTAGA TGTAACATT AAATATGTCA TTACAATCAT TTGCGTTAAA TATGTCACAT	5640
	CTGaTTCACC GTTATAGTGT TGtAAATTAC TATTTGTTAA AAAAGATATT GCTGTATTAA	5700
	ACGCTAAATC TATCGATTGG TTTAAATTAT GATTTGGATT TAAAAAAGC CATTGCTGAA	5760
45	CTATTAGCAA TACAAATGTT ATAAACCCCA TAAATCCATT AAATGCCAGA AAATGTTTGA	5820
	CATATGTTTT AGCTGACATG TGTCTAAAT CTGTGCCGAT AATTTTAAAA CACATATTTT	5880
50	CAAATCTAGT AAATATTAAA TCTACTCTTG ACGATTGCAC CAATGCTACG CGATATAGAT	5940
	ATCCACTAAA AACATACGTA ATCATAACCA TCATTGTTAG AAACAAAATT ATTTCCATGA	6000
55	TAACCCCTCAC TTAATATATT TCTAAAATTT TCACTACGA ATTAAGGCAT AAAATAAATA	6060

ACACAACAAC ATCGTAACAA CTTGTTTATG AGAGAAATnT TAATTTTCAA ACTTAGTTAT 6180  
 TAAGAAAnCA TTAAGATGTG TATGCAGAAA TAAATTTTAT AGCATTTAAT TGTGAAGAAT 6240  
 5 ATTATGATAT TGCTATCGAG GTGAAGGTTA TG 6272

## (2) INFORMATION FOR SEQ ID NO: 141:

(i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 1978 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

AAATGATGTT TTACAATAAA TATAnAAACG TATCAACATA TATCATCATA TTTTtagTTT 60  
 20 CAAGTGCAGC CTTTGCAATA TTCTTGTTAA GTGCGnACAT TAGTGCTCAC TCGGAACAAG 120  
 TGTACGAAAT GACTGACCAT CAAATTAAGA ACAATACGAT AAATAAAGCA TACGAACATA 180  
 AAGACCCTAC AAACAATAGC GAACAAAGAG ATGGGAAAGT GTTCGCTTTA ATAAATTGAT 240  
 25 ACATTGTCAC AACGTTATTT TGCCTATTTT TGCgmAATAG CGTTTTTTAT TACwTTTTTG 300  
 CTGATsTTAA ATTTGTTATA TTTTGTTAAA GTATTATAAT GATTGAATAA ACAAATTGAA 360  
 GGTAGGTTTT TTAATTGAGT AATTCTGATT TGAATATCGA AAGAATTAAAC GAGTTAGCTA 420  
 30 AAAAGAAAAA AGAAGTAGGA TTAACTCAAG AAGAAGCAAA GGAGCAAACA GCCTTAAGaA 480  
 AAGCTTATCT TGAGAGTTTT AGAAAAGGGT TTAAACAACA AATTGaAAAT ACTAAAGTAA 540  
 TTGATCCAGr AGGTAATGAT GTAACACCTG AAAAAATTAA AGAGATACAA CAAAAAGAG 600  
 35 ATAATAAAAA TTAAATCACA AATCTGTAAA GAATTTTCTG ACATTATAAC TTGAAATAAG 660  
 TATTTACTT ATCTTTTTAT TTAAAATAA GTTATAATGT ATTTGATAAA ATTGAAGAAG 720  
 GGAAGATACA CAAGATGTTT AATGAAAAAG ATCAATTAGC TGTTGATACG CTACGTGCAC 780  
 TAAGTATCGA CACAATCGAA AAAGCGAATT CTGGTCATCC AGGATTACCT ATGGGAGCTG 840  
 CCCCATGGC TTACACTTTG TGGACACGTC ATCTGAATTT TAATCCACAA TCTAAAGATT 900  
 45 ACTTCAATAG AGACCGTTTC GTATTATCTG CAGGGCATGG TTCAGCATTa TTGTATAGCT 960  
 TGTTACATGT TTCTGGTAGT TTAGAATTAG AAGAATTAAA GCAATTTAGA CAATGGGGTT 1020  
 CTAAACACC AGGTCATCCT GAATACAGAC ATACAGATGG TGTAGAAGTT ACTACGGGAC 1080  
 50 CACTTGGACA AGGTTTTGCT ATGTCAGTAG GATTAGCTTT ACAGAAGATC ACCTAGCAGG 1140  
 gAAATTTAAT AAAGAAGGAT ATAATGTTGT AGATCATTAC ACATATGTAT TAGCTtCTGA 1200

55

AAGTAAATTA GTTGTTTTAT ACGATTCAAA TGATATTTCA TTAGATGGCG AATTAAACAA 1320  
 AGCTTTTTCT GAAAACACAA AAGCTCGTTT TGAAGCATAT GGTGGAATT ACTTACTAGT 1380  
 5 TAAAGATGGT AATGATTTAG AAGAAATTGA TAAAGCGATT ACTACAGCTA AATCTCAAGA 1440  
 AGGACCAACG ATTATTGAAG TTAAAACAAC AATCGGATTT GGTCACCGA ATAAAGCAGG 1500  
 AACTAATGGT GTTCATGGGG CACCTTTAGG TGAAGTTGAA AGAAAATTAA CATTGAAAA 1560  
 10 TTACGGTTTA GATCCTGAAA AACGTTTAA TGTTCAGAA GAGGTATACG AAATTTTCCA 1620  
 AAATACTATG TTAAAACGTG CTAATGAAGA TGAATCTCAA TGAATTCAT TATTAGAAAA 1680  
 ATATGCAGAA ACATATCCTG AATTAGCAGA AGAATTTAAA TTAGCGATTA GTGGTAAATT 1740  
 15 GCCTAAAAAT TATAAGGATG AATTACCACG TTTTGAACGT GGTCATAATG GTGCATCTCG 1800  
 TGCTGATTCT GGTACTGTTA TTCAAGCAAT CAGTAAACT GTCCCTTCAT TCTTTGGTGG 1860  
 ATCAGCAGAC CTTGCTGGTT CAAACAAATC CAATGTAAAT GATGCAACTG ATTATAGTTC 1920  
 20 TGAAACACCT GAAGGtAAAA ATGTGTGGTT TGGTGTACGT GAATTGCTA TGGGTGCT 1978

## (2) INFORMATION FOR SEQ ID NO: 142:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7588 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

TAGTAGTATT TATTAAATTA TACGAAGGGA CCcAACACAG AAAATTCATT TTATTGAATT 60  
 35 TTACATTTAT GTGCCAAGTT GGGAAAAATG TCTTATTTT TCaAAGTATT TAAAAGTAAA 120  
 ATTACATGTT AATACGTAGT ATTAATGGCG AGACTCCTGA GGGAGCAGTG CCAGTCGAAG 180  
 40 ACCGAGGCTG AGACGGCACC CTAGGAAAGC GAAGCCATTC AATACGAAGT ATTGTATAAA 240  
 TAGAGAACAG CAGTAAGATA TTTTCTAATT GAAAATTATC TTAGTGCTGT TTTTtaggga 300  
 TTTATGTCCC AACCTTTTAA GAATATTAAA TTTCTACAAT TTCGTCATCT TCAACAATAA 360  
 45 AGCCCATTGT ATTGACGCTG TTATTTAAGA AAGTCAGAAT ATAACGCATT ACTTCATCAC 420  
 GTTCTGGCTC ATTGTGAACC TCGTGGTAAA AACCTTGCCA AGCTTTAAAA TATAATTCAG 480  
 GTGTTTGATA TTTTCTTTA AACTCATCAA TTGCCCTAGT ATCAACAATT AAATCCTTCG 540  
 50 TTCCATACAT TAATAGCGTT GGCATTGGTT GAATGTCATG AATATGAGCC ATCGTATCTT 600  
 TCATCGTCTC ATTAATTGTA TTATACCAAT GATACGTTGC TTTTTTTAAC ATTAAACCAT 660

55

	CATTAAAACG TGTGTCITTT GAAATTTTAC CTATATTTGA AACAAAGTTTA TCTTTACGAT	780
	TTTTTCCATT CTTTGAAGT TCTAGCATAG GAGAAATTAA CATCATCCCC TCGATTGGCA	840
5	ATTCTACTTT TTCAAGTAAA TTTAATAAAA TCAAACCGCC AAGTCCTACC CCTAATACAT	900
	AAGTAGGAAT TTTATATTCA TTAGCTATCT TTAACCAGTC TAGCAAACCTT TCGTGATACG	960
	TTTGAAAGTT TTCAATTTGT CCTTTATTAG CTCTTGAAGT TTGACCTTGA CCAGGCAAAT	1020
10	CTCCCATAAAT CACATGATAG CCATTTCTTC TTAACATCGT AATAACATAT GCATATCTTC	1080
	CCGTATGTTT TAATATATTA TGAGCAATAA CAACGACGCC TTTCGCATCA TTTTCAGCTT	1140
	CCCACCTCCA CATTATTATA CTGCCCCCTT TTCATTAATC TTCAATAACA TAATTATAGC	1200
15	AAATTCACCTA TGTTAGATTTT TATTTATAGT ATTATTGTTG TCCATATTAT TATATATAAA	1260
	TGAAATCAAC ATCAATAATA GTGTAATTAT ACATAATTAT TTTTGATTGT TTTTGATGAA	1320
20	AACGCTTTCT CGAATATTTT TTTTCATGCTA AACTTATTGT AAACACAAGG GTTTGGAGGA	1380
	GTAGCAATGG CACTATTAAA GAATTTTTTT ATCGGATTAT CTAATAATAG TTTTTTAAAC	1440
	AACGCAGCAA AAAAAGTGGG CCCACGTTTG GCGCCCAATA AAGTCGTTGC CGGAAATACA	1500
25	ATTCCAGAGT TAATTAATAC AATCGAATAC TTAAATGACA AGAATATCGC TGTTACGGTA	1560
	GACAATTTAG GGAATTTGT CGGTACAGTT GAAGAAAGTA ATCATGCTAA AGAACAAATT	1620
	TTAACAATTA TGGACGCGCT TCATCAACAT GCGGTAAAGG CACATATGTC TGTTAAATTG	1680
30	AGTCAGTTAG GTGCAGAATT CGACTTAGAA TTAGCTTACC AAAATTTAAG AGAGATTTTA	1740
	CTTAAAGCAA ATACTTACAA CAATATGCAT ATAAATATTG ATACTGAAAA ATATGCTAGC	1800
	CTGCAACAAA TTGTTCAAGT TTTAGATCGC TTAAAGGCG AATTTAGAAA TGTTGGTACT	1860
35	GTAATTCAAG CATATTTATA CGATAGCCAC GAATTAGTTG ATAAGTACCA AGATTTACGA	1920
	TTACGTTTGG TTAAAGGTGC ATATAAAGAA AACGAATCAA TTGCATTTCA ATCTAAGGAA	1980
40	GACGTAGATG CAAATTACAT CAAAATAATT GAACAACGTT TGTTAAACGC ACGCAATTTT	2040
	ACTTCAATTG CAACACATGA CCATCGCATC ATTAATCATG TAAAACAATT TATGAAAGAA	2100
	AATCACATTG AAAAAGATCG TATGGAATTC CAAATGCTCT ATGGTTTTAG ATCAGAGTTA	2160
45	GCAGAAGAAA TCGCAAATGA AGGCTATAAT TTCCTATTTT ATGTACCTTA TGGCGATGAT	2220
	TGGTTTGCGT ATTTTATGAG AAGATTAGCA GAACGCCAC AAAACCTATC TCTTGCTGTA	2280
	AAAGAATTG TGAAACCTGC TGGCTTAAAA CGTGTGGCA TAATTGCAGC TTTAGGAGCT	2340
50	ACAGTTATGT TAGGTTTAAAG TACAATTAAA AAATTATGCC GTAAATAGAG CAAGACATAA	2400
	ACAATAATTT AGGAGTCTGG AACAAATATC AATGTTCTAG GCTCCTAAAT GTTATATTGG	2460

55

TAGATTTTAA TAAATTAGCC ATTTCAATTG CACTTACTGC TGCTTCAGCA CCTTTATTGC 2580  
 CAGCTTTCGT ACCTGCTCTT TCCACAGCTT GTTCAATAcT TTCAGTCGTT AAAATACCAA 2640  
 5 ATATGACTGG TACATTAGTT TGATCATTCA CTTTAGAAAC ACCTTTCGCG ACTTCATTAC 2700  
 AAACATAATC ATAATGAGAC GTAGCACCGC GAATTACGCA TCCTAATGTA ATTACTGCAT 2760  
 CATAATTTCC TGATGAGGCT AATTTTTTAG CTAATAAGG AATTTCAAAC GCACCTGGCA 2820  
 10 CAAATGCTAC ATCAATATTG TCTTCATTAA CATCATGTCG AATCAAAGTA TCTTTTGCAC 2880  
 CTTCAAGTAA TCTTCCAGTG ATAAAATCAT TAAATCGACT AACTACGATT GCAACTTTCA 2940  
 AATCTTTTCC AATTAATTTA CCTTCAAAAT TCATGTTAAA ATCCTCCTAT ATTAAATGAC 3000  
 15 CCATTTTAT TTTTTTCGTT TCCATATAAT CATGATTATG TACCGTTTCT GGTACGATAA 3060  
 CTTCAATTCT TTCTGCAATA TCAATGCCAT ATTGTTTTAA TCCCTCAAAT TTACTTGGAT 3120  
 20 TATTACTTAA TAAATTGATA TGTTGATGT TAAAATATTT TAAAATCTGT GCAGCAATAT 3180  
 GATAATCTCG CAAATCTTCA TCAAAACCTA ATGCTAAATT TGCAGTTACT GTATCATATC 3240  
 CTTGCTCAAT TAATTCATAT GCGCGTAATT TGTTTAACAA TCCTATGCCA CGACCTTCTT 3300  
 25 GAGGTAGATA AATAATCATG CCACCATGTT CATTGATATA CTTCATAGAC GATTCAAGTT 3360  
 GAGCACCACA ATCACAACGT TGAATATGGA AAATATCGCC TGTAAGgCAC GCAGAATGTA 3420  
 AGCGTACATT TTCATGTTGT CGAATTGCAC CTTTTGTCAG TACAACTATC TCTTCATCTG 3480  
 30 TGTATGTCGCG TTTAAACCA TACATATCAA ATGTTCCGAA ATCTGTAGGC ATTTTCACTT 3540  
 TTGCCTTAAA TTCAATTTCT GGTTC TAATT TTTTACGATA TTCAATTAAA TCATCAATCG 3600  
 TAATCATCTT TAATTGATGT TTTTCTTTAA ACTTTTGTA ATCTTGTCCT TTCGCCATCG 3660  
 35 TGCCGTCATC ATTCATAATC TCACAAATGA CACCAGCGGG CTTGGCACCA GTAAGTTTAG 3720  
 CTAAATCAAC AGCCGCTTCT GTGTGTCCAT TTCTAGCTAA TACGCCTTTA TCTGTGCTA 3780  
 CTAATGGAAA TAAATGACCA GGACGATTAA AATCTTTAGC TTCACTACTA GGATCAATGA 3840  
 40 GCTTTTTGGC AGTCAATGTA CGTTCATAAG CACTAATTCC TGTTGTTGTA TCTACATGAT 3900  
 CAATACTCAC TGTAATGTC GTACCAAAGA TGTCGGAGTT ATCATCAACC ATTTGTACCA 3960  
 AATCCAAACG TTGTGCAATA TCTTTAGACA CTGGTGCGCA TATTAATCCC CtTGCTTCTT 4020  
 TCGCCATAAA ATTAATGGTA TTATCGTTCA TCCATTAGT AACCGTACT AAATCACCTT 4080  
 CATTTTCACG ATTCTCATCA TCTACTACAA TAATTGGTTC TCCATTTTTT AAAGCCATTA 4140  
 50 AAGCACTGTC AATATTATCG AATTGCATGC TACCCCTCct AAAAACCAAA TGCTCTTAAT 4200  
 TTATCTACAG ATAATTGGTC TTTATCTTTA TTTAAATAT TTTCAACATA TTTAAACAAA 4260

55

	CTCGTTTCTG GAATAAGATG AATGTCAAAA CTGTTATCAT GCTTATCAAA TACCGTTAGA	4380
	CTAACACCAT CCACAGTAAT AGACCCTTGC TTAACCTAAT GATTATTAAT ATGTTGGCTA	4440
5	CATTGAATCG TAATAATTTT TGCATTGGCT GTTTCATTTA TTTTGAAC TGTTCCTAGT	4500
	TCATCTACAT GACCGAGGAC AAAATGTCCA CCAAACCTAC CGTTACCACT CATGGCACGC	4560
	TCTAAATTTA CTTCTGATTG TCGCTTAACA TCTGCTAAAT AGGTTTTATT TTCAGTGCCT	4620
10	TTAATTACTT GAACAGTAAA AGATGTCTGA TTAAATCAA TCACTGTAA ACATGCACCA	4680
	TTAACTACTGA TGAATCACC AATATGCATA TCTGCCGTAA TCTTATGTGC TTCAATTTCA	4740
15	ATCGTCCTGA CTGATTGACG AATTTGAACA CTTTAAACGA CACCTATTTT TTCAACGATG	4800
	CCAGTAAACA TGCATCATCA CTTCTTTTCT AAAGTTAATT TAACATTTTG ATTTAATAAC	4860
	TCGGAATGAA CAATTTCAA TTGGTTCGCA TCTGGTATCT CAATCACATC ATTTGTTTGA	4920
20	TAAAATTGAT AATTTCCAGA TCCGCCAATT AATTTGGGG CATAATAGAG AATAAATTCA	4980
	TCTATATAAT TAGATTGGAG AAATTCTGAA GTAGTGGTTG GACCTGCCTC GACTAGCAAA	5040
	GTTCCAACCTC CTCTTTTATA TAAATTGTGA AGAATTGTTG TTAAATCGCA AGACTTCAAG	5100
25	TAAATAATTT CAATATGTGT TTGATTGGTT GTTAAATTTG GATTTTCAGT ATATATCCAA	5160
	ATTGGTGTG ATTCACTCTG ATAAATTTGC TGATTAAAT GAATATTCCC AGACTTAGAC	5220
	AATATTACTT TTATAGGGTT TTTTCCATCT TGAATACGTG TAGTATATTG TGGATCATCT	5280
30	AATCAACTG TACGTCTTCC AGTTAACT GCGTCGTGTC GATGTCTTAA CTTATAGACA	5340
	TCTTGTTTAA CCTCTTTGTT AGTAATCCAT TGAATTTGTC CATTATCATT CGCTTGTTTA	5400
	CCATCTAAAC TTGCAGATAC TTTCCTGTG ATTTGTGGCA GTTGCTTTGC TTTTGCTTTA	5460
35	AAAAAGTCTT GGTATAATTG TGATGCCCGT TCATCATCAA CGCAATCAAC CTCAATACCG	5520
	TGAGCCCGTA ACGTCTCATC ACCATGTGTG TCTAACGAAT TGTCTTTTGT TGCGTATACT	5580
40	ACTTTTGCTA TCTTACAATC AATTATTTTG TTAACACAGG GTGGTGTGA ACCAAAATGA	5640
	CTACATGGCT CTAACGTAAT ATAAATCGTC GCACCTTCAG CATTTTGTG TGCCATATCA	5700
	AGTGCTTGAA CCTCCGCATG CTTGTACCT TTTCTCAAGT GTGCACCAAT ACCAACAATC	5760
45	CTACCTTCTT TAACTACAAC AGCGCCAACG GGTGGATTAA CACCTGTTTG ACCTTGTAAC	5820
	ATATTGCAA GTTGAATCGC ATAATCCATA AATTGACTCA AATGATCACC TCTATAAACA	5880
	AAAATCCTCA CATCATGAAT TAAGATGCAA GGAGAAAAAT TTATCGTTAA ATAAGCCTAT	5940
50	TTGTACACAT TTTTACAAAT ACGCTACATT ATCTTTGTCTG ATAATTAACA TTCTTTCTCC	6000
	CATCCAGACT TTAAGTGTG GCTCTAGAAT CTCCTAGAT CAGCCACTAA TATGAAACAT	6060
55		

TTaTATATGA AATTGTTATA GATTATTTGA GTACGTAGTA TGTCAACTAC ATTTAAAATG 6180  
 ATACTATATG TTTTCTGAAA AAACAATTAA TGACGGTTTT AATTTAATAT AATCTGAGTA 6240  
 5 CTATAGGCAT CTCATTGATA TGATTCTTAC TAACAGACAT TAAAATCAAA CCTTCAATTC 6300  
 GTCTCTATAG AGCGTTCTCT TTATTATCTT CTAGTTACAA ATTATTGATT GtCACtGCGC 6360  
 TGTGTGTGCT CATTGATTTC TAAAGCATCA TATAATTGAG ATACTGTATG CGCAACTTGT 6420  
 10 TCTACAATCA TTTTCACACC GTTTCGTAGT TTATTAACAC CGTTTGTCTT TTGACCTATC 6480  
 GCAATCATAT TTGTTAATGT TCCAAACCTT GGACTAATAA CTTGATTGGT TTCCGGAATG 6540  
 ATTTGTATGC CTCCATTGG GTGTGCTTGT ACAATTTGTC TATTTTCAAG ATTTCTAATT 6600  
 15 AATTGATCAT CTTGATCCAA TTCATTTAAA TGACTTTTTG CACCTGTGCG GTTAATGACA 6660  
 ACATTATATA TGTCTACTGA TTCTTGGTTT TTGTATGAAA AATAATACAA CTTGCCATaC 6720  
 ATGTTCCACAT CTTCTAAATC TTTTTTCAAA ATTAAAGACT TATTTTCTAT TAATTCAATA 6780  
 20 ATTAGTTCAG CAGTCTTGG AGGCATTGGA TTTGAATTTA ATTGAATCAT CTTTGAGTAT 6840  
 TTTTGATTAA ATTGATGTTG GTCTTCAATA CTTAAGCTAT TCCATATCCA ATTTAAATTC 6900  
 25 TCTTTCAAAT GTTCAATCAT ACTTTGAAA ATGCCCaTTT CTGTTGGACG CGCTAAATCA 6960  
 TACTTCAAAT CTGCAATATG ATTTCTGTG CGTCTATGTA CTAATTTTTT AAAATCAATG 7020  
 TCATATTCAG CACATTCTTT TAAAAATAAA GAAACTAAAG TATCAAGCGG TGCATTGCCG 7080  
 30 AAATGATGTT TTTTAATGTC ATTTAATTTG TCTTTAGTTA AGTACTTGAA TGTACGCTCT 7140  
 ATCATTGTAC CTCTTACACT TGGTAAATGA GCAGAACGAC TCGTCATAGT AATTGGTAAT 7200  
 TTTGGATGAT GAGCAGCAAC ATAACGGACA ACATCTAAAC TGGCAAGGCC TGTACCAATA 7260  
 35 ATCGCAATAT CGTCCAGTTC ATTTACTTCG TCTAACGTAT TATATGTTGG ATAAGGCGTA 7320  
 gcGATATATC CTTTTTTACC CTTTAAGTTA TATGGATCAT GGTAGGCAAA TGTACCACAT 7380  
 GTTAAAAATA CATAATCGTA CGCTTGCCAT GATTGTCCTG AATTTGTAGT ACATATGTAA 7440  
 40 TAAGTTAAAT TCGTTTCATC GATATTAGAA TTTGTATAAA TCTCTGAAC TTTATTATAA 7500  
 TTAGTTGATA TATTTGGATA TTTTTCGTG AACATAGATA AATAAGATTT CATATAATGT 7560  
 45 CCGAATACAA ATCTCGGTAA ATATGCAG 7588

(2) INFORMATION FOR SEQ ID NO: 143:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10320 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

	nCTAGGTATT TTAAACCTAA TCTAGATAAA CTAGCTTCGT AAGCAGCTGC TACATTTTCA	60
5	CGACCGAAAT CCTCAAAATA TAATTTTGAA GTAATAAATA AGTCTTCTCT AGCAATACCA	120
	GTTGACTCCA ATCCGGCACG AATGCCAGCA CCTACTTGTT CTTCAATCCC ATAAACTTTT	180
	GCGGTATCAA TACTACGATA TCCTTGTTCA ATGGCATACT TAACACTTTC CATGCAATTT	240
10	TCATCATTTT CCACACGAAA TGTCCTAAA CCAATTTGTG GCATCGTGTT TCCATTATAA	300
	AATGTTTTAA CCTCCATAAA TATCGCCTCA CCTTTTGTGAT GTATTATACC CTGTTATCAT	360
	AACAAATCTG AGTTGAATAC ATGAGAAAAA AACTTAGAG CAATCAACCA CTAAAATTCT	420
15	AGTAATATCT CTCAAATATT AATCAAATTG TAAAAGTAAT TCTGTTTAAT TTATGACAAA	480
	CTAAAAAGC CGAAGTAACA ACATATAGTC ATCACTTCAG CCTAACATTT AATTGAATGA	540
20	TTCAATTTTA TCCATCATTT GTTGTAAGTC TTCCACGTTG TATTGAATAC GACCATGGAA	600
	TACAAATTTG TTAAAGAACT CGTCTAATTG TTCAGCACCG ACAAGCACTT TGACAGCACT	660
	ATTTTGATTA TAATTGAAA TCGTTACATC GCCTTCATTT TTAAGATTAA AGTATAAAAT	720
25	TGAAGTTGGT GTATATTTGG CACCTAATTC TTTTGTAG TCTTCAGCCA ATTGTTTAAT	780
	CGCCTCAATT TGATCTGAAT AATTTACAAA TGATAATGAA CGTTTGTCAAT CATTTTGATC	840
	CATCACAATA GTTTGCGGTC TAGATTTATC TAAATCCAAT GTATCAAATA CTTGTTCCAT	900
30	TGGTGGTAAA TCTTTAAATT GACCGCCACT AATACCATT AAAACATGAC CTTTAAACAA	960
	TTGAGAATCA ATAATATAAA GACCAGTTCT TGTTAATACT AAATGACTAA TTCGTTCAAT	1020
	ATTATTAAAG CCATCCTTTG GTAAAAAGAT ATTTGCCATA ATGTGCATAT CTTCTGGTCG	1080
35	AATTCGTTTT TCTTTAACTA ATCTTTCACG AATACCAATT AATCTCATGT CCGTTACATA	1140
	TTCACTATGA TTTTTCGAGA ACAATTTTAA TGCGTCAATC TCACGATCTT TTGTACTAAC	1200
40	CATGTGATTA TAATCTTCTT GTTGTTTTGT AATTGTCTTT TTATTTTGAA TACGCTCTTT	1260
	CTCTAAAGCT TCTTCATGAG ACTTTTAAAT GTTTTGTCTT TGTGTTTCAT ACTTTTCTTC	1320
	TGTTTGTGCG TTAACTTTTT TCTTACTACC TAAGGCAACT AAAAAAAGGA CAAAAAGAT	1380
45	TAATGCAATG AgCTACTGCA ATAATGAGTC CAATGACTAT CGGTGAAGAT AAATCCATCA	1440
	CAACAACGCT CCTTTTAAAT ATATGAATAA CTTTAATTAT AATAGAAAG CTAAAGATTT	1500
	TCGATACATA TTATCATTTA TATACCGAAA ATCTTTTATT TAGCTATATT CAATTCATCT	1560
50	TATTATTTTA CTGCGTCTTT TAATCTTCC ACTTGTCTA ATTTTCCCA TGGGAATAAG	1620
	ACATCTGTAC GTCCAAAATG ACCATAAGCA GCAGTTTGTT TGTAATCGG TTGTTTCAAA	1680

55



	AGTTGCCCTT CAGAACTTT ACCTGTTCCA AATGTATCAA TTGCAATTGA CACTGGTTCT	1800
	GCAACACCAA TCGCATATGC CAATTGTA CTACATTGAT CTGCTAAACC TGCTGCAACA	1860
5	ATATTTT TAG CCACATAACG TGCAGCGTAT GCAGCTGAAC GGTCTACTTT TGTAGGATCC	1920
	TTACCACTGA AGCATCCGCC ACCATGACGT GCATAGCCAC CGTACGTATC AACAAATGATT	1980
	TTACGTCCTG TTAATCCTGC ATCACCTTGA GGTCCACCGA TTACAAAGCG TCCTGTAGGA	2040
10	TTGATGTAGA ATTTAGTTTG TTCATTAATC AAGTTTTCTG GAACAGTTGG ATAAATGACA	2100
	TGTGCTTTAA TGTCTTCTTG AATTGTTC AAGTGTACAT CCTCAGCATG TTGTGTTGAT	2160
15	ACGACAATCG TATCAATACG TACTGGGTTA TCATTTTCAT CATATTCAAC AGTGACCTGA	2220
	ACTTTACCGT CTGGTCGTAA ATAATTTAAC GTACCATCTT TACGCACATC TGATAAACGT	2280
	TTTGCCAATT GATGTGATAA ATAAATTGCT AGAGGCATAT ACGTCTCTGT TTCATTCGTT	2340
20	GCGTAACCAA ACATTAAACC TTGGTCACCT GCACCTGTTG CTTCAATTC TTCTTCGCTA	2400
	TCTTTATCAC GATACTCTAA TGCTTTATCC ACGCCTTG TGCAATGTCAGG TGATTGTTCA	2460
	TCAATCGCAG TTAATAATGC CATTGTTTCA TAATCATAAC CATATTTTGC TCTTGTGTAT	2520
25	CCAATTTCTT TAATTGTTTC TCTAACAACT TTCGGAATAT CAACATATGT TGTGTAGAA	2580
	ATTCGCCCGG CGATCAATGC CATACCTGTT GTAACAGTTG TTtCACAAGC TACACGTGCA	2640
	TTTGGATCGT CTTTTAAAT AGCATCTAAT ATTGCATCTG ACACTTGGTC AGCGATTTTA	2700
30	TCTGGGTGTC CTTCTGTAAC AGACTCTGAA GTAAATAATC GTTTGTTATT TAACATAGTT	2760
	TGCTCCTTTA AATTTATATT ACGAAAATTC TCTCTCTGTG AGCTAAATAA AAAAGACCTT	2820
35	CTAACTATTA ATATAGAGAG AAGGCCTAAT ACGTCCATTC GCTCTTATCG TTCAGACCTA	2880
	TTTGTCTGCA AAcGGTTTGG CACCTTTCTT TTATAAAAAA GAGGTTGCTG GGTTCATTG	2940
	GGTCCATGTC CCTCCACCAC TCAGGATAAG AGAATCCGTT AAAAATAATA GTACCTAATT	3000
40	AATGAATTAA TGTCAATTTT TCACAAATAA ATTTACAGTA AAATATTGTA GATTAATTAT	3060
	GTTAATGTGT TATACTAATT AAATGTAAAG GCTTACATTT AAATTATCGC TTTGGAGGGA	3120
	TTTAGGATGT CAGTAGACAC ATACACTGAA ACAACTAAAA TTGACAAATT ACTGAAAAAA	3180
45	CCAACGTCAC ATTTTCAACT TTCGACGACA CAACTTTATA ATAAAATCTT AGACAATAAC	3240
	GAAGGGGTAT TAACAGAACT TGGTGCTGTT AATGCAAGTA CTGGAAAATA TACTGGTCGT	3300
	TCGCCTAAAG ACAAATTTTT TGTCTCTGAA CCTTCATATA GAGATAACAT TGATTGGGGA	3360
50	GAAATTAATC AACCTATCGA TGAAGAACT TTCTTGAAGT TATACCATAA AGTACTAGAC	3420
	TATTTAGATA AAAAAGATGA ACTATACGTA TTTAAAgCt ACGCTGGTAG CGATAAAGAT	3480

55

	ATGTTTATTA GACCTGAATC AAAAGAAGAA GCTACAAAGA TTAAACCTAA CTTCACTATC	3600
	GTCTCTGCAC CACATTTTAA AGCAGATCCA GAAGTTGATG GTACTAAATC TGAAACCTTT	3660
5	GTCATTATTT CATTTAAACA CAAAGTCATT TTAATCGGCG GTACTGAATA CGCTGGTGAA	3720
	ATGAAAAAAG GTATCTTCTC TGTAATGAAT TATCTCTTAC CGATGCAAGA TATTATGAGC	3780
	ATGCATTGCT CAGCAAACGT TGGTGAAAAA GCGGATGTTG CATTATTCTT TGGTCTATCT	3840
10	GGCACTGGTA AAACAACCTT ATCGGCTGAC CCACACCGTA AACTAATCGG TGATGATGAA	3900
	CACGGCTGGA ATAAAAACGG GGTCTTTAAT ATCGAAGGTG GCTGCTATGC AAAAGCAATT	3960
	AATCTTTCCA AAGAAAAAGA ACCACAGATT TTTGACGCAA TCAAATATGG TGCAATTTTA	4020
15	GAGAACTG TAGTTGCAGA AGATGGTTCA GTGGACTTTG AAGACAATCG TTATACAGAA	4080
	AACACGCGTG CCGCTTATCC AATTAATCAC ATTGACAATA TTGTAGTACC ATCTAAAGCA	4140
20	GCACATCCAA ATACAATTAT TTTCTTAACT GCGGATGCAT TTGGTGTAT TCCACCGATT	4200
	TCAAAGTTAA ATAAAGACCA AGCAATGTAT CATTTCTTGA GTGGTTTCAC TTCTAAATTA	4260
	GCTGGTACAa GCGTGGTGTG ACAGAACCTG AACCATCATT CTCAACATGT TTCGGAGCAC	4320
25	CGTCTTCCC GTTACACCCT ACTGTTTACG CTGATCTATT AGGTGAACTT ATCGATTTAC	4380
	ATGATGTTGA TGTTTATCTT GTTAATACTG GATGGACTGG CGGAAAATAT GGTGTAGGAC	4440
	GTAGAATCAG CTTACATTAC ACACGTCAA TGGTAAACCA AGCGATTCTT GGCAAATTGA	4500
30	AAAATGCAGA ATATACAAA GATAGTACGT TTGGTTTAA GATTCTCTGTA GAAATTGAAG	4560
	ATGTACCGAA AACAATTTTA AATCCAATTA ATGCTTGAG CGACAAAGAG AAATATAAAG	4620
	CACAAGCAGA AGATTTAATT CAACGTTTTG AAAAGAACTT CGAAAAATTT GGTGAAAAAG	4680
35	TTGAACATAT TGCTGAAAAA GGTAGCTTCA ACAAATAAAT TTGAATACTA AATCaAAACC	4740
	ACCGGTGTGA ACGGGTGGTT TGTCTGCGG CTATAAGCCT TCCTTACTGG CCAGCCCTAA	4800
40	AAGGGCACTG ACAAGTCAGC CAACTGCACT ACTATTCCAG CAACCCTAAA GGGTTACTCT	4860
	TTTTCTTTC TTTTTTTATT TTTCTCTCCA GTGAAAGGAT CTAAATATTC TTCCATTGAG	4920
	ATTTGGTCTG CAACGATATC CTCTTGTAAT TGATTACGAA TATAATTTTC AATCACTTTT	4980
45	TTATTTCTAC CTACTGTATC CACATAAAAT CCTTTACACC AAAACTTTCT ATTTCCATAT	5040
	CTATACTTTA AGTTAGCATG TCTATCAAAT ATCATTAAAC TACTTTTTCC TTTTAAATAG	5100
	CCAACAAATG ATGATACCCC AAGTTTGGGT GGTATACTAA CTAACATATG GATATGATCT	5160
50	TTACATGCCT CTGCTCAAT TATCTCTACA CCTTTTCTTT CACATAATTG ACGCAATATA	5220
	ATCCCTATAT CTTTTTTTAT TTTTCCATAT ATCACTTGTC TTCTGTATTT AGGTGCAAAG	5280

55

	AAATAGCATC TCCTCGTGTT GATTATTTTG GTTGGCTGAC CAATATTTAT TCTAGCACGT	5400
	AGAGATGCAT TTTTGTGAC AATGGTAGAA CCTTTTctGa ACCATACGCA TAGCGTATGG	5460
5	TTTTCTTTTT ACAATTAAAG AGCCAACCGT TGTATAGTC TAACAATGGT TGGCTCCTCT	5520
	TATTTTATGT GCTAAAAATT TATAGGCAAT TTTATTACAA CAATGTACAT TTAAGGTGAC	5580
	CTTCATGCCA AAATCGCATC ACTCATTTAA TGGGAAGCAGC ACGTCTTCAT ATAAAGTACC	5640
10	GATCCCTAAT TCAACGCATG TAGTACCACA TCTTCAAAGC TTGATAGTTC CCATGCGCAC	5700
	ACCACGTTTC ATACTAGCTA TCGGACTCAA CTGCGTTCAT AAACCTTTTA ATATAAGTCA	5760
15	ATGTTTCAAC CATCGCTGGT GGTCTTGGCA CATGTCCTTC TGCCATTTGA TAAAATGTTT	5820
	CATGCGTGGC ACCTTTTAAC TCTAGTTGGT CCGCTAAATA ATACGCATGA TGAATACCAA	5880
	CTTGCTGGTC TTTCCCTCCA TGTACAATTA ATATGGCGG ACTGTTTTCA TTAATGTTTG	5940
20	GAATCGCTTG GCGTGCCTCA TATGCCGCTC GATCTTTTTT CGGATGACCA ATCATTCTTC	6000
	GTAGCATGCC TCTTAAATCG ACACGTTCTT CATACATTAA ATCAATATCT GAGACACCAC	6060
	CCCAGATTGT ATAACITGTT ACTGGTAAGT CTTGAAATGT CAACAATCCT TGTAACCAC	6120
25	CTCGCGAAAA ACCAACCATG TGGATAAATG CATGTGGATA TTTATCATGT AGCAACCTTA	6180
	ATAATTGCGT CACATCATTT AAATCGCCAC GGTAAAATTC GTCTTTGCCT TCACTCCCAT	6240
	TGTTACCTCG GTAGTATGGC CCAATCACTA AAGTTTGA CTCTGAAAAT TGCATTAATC	6300
30	TACCTGCGCG CACACGTCCT ACTTGACCTT TGCCACCTCG CAAATAAACT ACAATGCGAT	6360
	TTACTTCATG ATGTGGTGTC ATCATTAAAG CTTTTACTTG TAAGTCATCT GACAAATATG	6420
	TAATTTCTTC GAATTGATGC GTAAAATATT CAATTGGCAT TCGTTTACGT TTGATAAAAC	6480
35	CCAAGTGATT GCACCCTCTC TACGCATTTT AAAATGGTAC TATCTTGACG TAAGAACTC	6540
	CGTTGTGCGA GTTCAATATC ATTGATACAG TTAAACAACA CTGGCCCTGC TGTCTCTAAA	6600
40	TAATCGTTCT TGCTTACCAA TGATTCAACT TCGATAAAAT ATACATCTTT TACAAAATCA	6660
	GTTTGATCAT GTGTTTCAAT GGTATATTGT GCTATGTAAT AAATATTTTT AACTTTGGCG	6720
	CCTGTTTCTT CATATAATTC aCGTGTA ACTGCTCAGCAC TACTTTCCCC GCGTTCCCTT	6780
45	TTACCACCAG GAAATTCAAT CCCCCGTAAA TTATGTTTGG TAAAAAGCAA TTGATTTTTA	6840
	AACGTTGGAA TAGCTAGCAC ATGATTGCCA TCTGCTATCT CATTATCCTT TTAAATGTC	6900
	AAATTAACTT GACGATTATC TTTATCCCTA AACTTCACGC GCATCACATC CCTACATTGT	6960
50	ATGTTAATAT AATAGTTAAT TACTATCGTT GGAGGCATTA ATTATGAAAA AGATATTCTT	7020
	GGCGATGATT CATTTTTATC AACGTTTCAT TTCGCCACTC ACTCCACCAA CTGTGCTTTT	7080
55		

	CCTTTATTTA	GGTATCCGTC	GTATTTTAAA	ATGTCATCCG	CTTCATAAAG	GCGGCTTTGA	7200
	CCCTGTTCCG	TTAAAAAAG	ACAAGTCAGC	AAGCAAGCAT	TCACATAAAC	ATAACCATTA	7260
5	ATATGGTTGT	AATTGAGTTA	TATCCACTAA	AGGGGGGCGA	AATTCGAGTC	GCCCCCTCTTT	7320
	TAATATGCCT	GAATGCGCCA	CCACATCTTG	TTCAAAATAA	TAACCTGCTG	GTGTAACATC	7380
	TCCTGGATAA	TCACCTTTAC	GAGCAAGCAT	CGCTGTAAAA	TAGCGGCTTA	AACCATATTC	7440
10	GTACATGCCG	CCAATAACCA	CTTTTGCAAC	ATGACTTTTC	AAAGTATCAA	TGCGCGTTTG	7500
	CACCTTTATCA	ATGCCACCTA	GACGAAATGG	TTTTAATACA	ACAACTTTCA	CATTGTATAA	7560
	TTCTATCAAA	TTAATTATGT	CCaACAACGA	TGTTGCCTTT	TCATCAAGGG	CTATTGGAGG	7620
15	TATTGTTCCA	TCCGCTACTT	CATCAAGCAT	GGAGATATCT	TTAAATGGCT	CTTCGATATA	7680
	AAGAACCTGT	TCACGCGCTA	ATAACTGTAA	CTGTGTGAAA	TCTTGACGAT	CCAAGGACTC	7740
20	ATTTGCATCT	ATAACCAATT	GAAAGTGAAA	GTCTAATTCC	CGTAACACTC	TAATTTGATG	7800
	CATGATTTGA	GGCGTCCATT	TTAATTTAAT	TCTGGTCGGC	TTTGTGTGCTT	TTAATGACTC	7860
	TAGTTGTTTA	TTTGATAAGC	CGCTCGcTGT	CGCTCCATAT	GCTACTGAAA	ATGAAGGCAG	7920
25	TACATGAAAC	ATTTGATACA	ATGCCATGAC	AATAGTTGCC	CTTGCAGCAG	GCGTATTTTC	7980
	CAATGAATCT	ACTAATTTTA	GTGCTGCTTC	ATACGTTTCA	AATGATTTAT	TTCTATTATC	8040
	TTCGAACCAT	TGCTCAATTA	CATGTTTCAC	TGAGGCAATT	GTTTCATGAT	CATACCAATC	8100
30	TGTTTGAAAA	GCGTTACATT	CCCCGAAATA	TGCATTTCTT	TTGTCATCAA	TCAATTGCAT	8160
	AAACAAACAA	TCACGATGCG	TTAAAGTGAC	TTTCGGTGTT	ACAATTTGTG	ACTTAAATGG	8220
	CTCACTATAT	TTATAAAAAT	GCAAAGCTGT	CAACTTCATC	AAATCATCCT	CTATACAAC	8280
35	TATTTCTTTG	TAATTTACCT	GTTGATGTAT	AAGGTAAAGT	ATCAACCTTT	TCAAAGTGTT	8340
	TCGGTACTTT	ATATTTGCT	AAATGTTGTG	ATAAATATGC	AATCAATTGT	GCCTTTGAAA	8400
40	TGTCACCTTC	ACTGACAAAA	TATAATTTAG	GCACTTGGCC	CCAAGTATCA	TCAGGATGCC	8460
	CTACACATAC	TGCGTCACTG	ATACCTGGAA	ATTGctTCGC	TACCGTTTCA	ATTGATATG	8520
	GATAAATATT	TTACCCGCCA	CTAATAATTA	AATCTTTACG	TCGGTCATAA	ATCATGACAT	8580
45	AACCTTCATG	ATCTATTTCA	GCAATGTCAC	CCGTATTAAA	ATAACCATTT	TCAAACGTAC	8640
	CCGTAAATC	TGTTGGATAC	AAATATACAT	TCATCACATT	GGCGCCTTTA	ATCATTAATT	8700
	CTCCATGACC	TTCTTTATTA	GGATTTTAA	TTTTTACGTC	AACATTGGCA	CTTGGCATCC	8760
50	CTACAGTGTC	AGGACGTGCA	TGCAACATTT	CCGGTGTTGC	TGTTAAAAAT	TGCGAACATG	8820
	TCTCAGTCAT	ACCAAATGAA	TTATAAATTG	GCAGGTATA	TTGTAATGCC	GTCTCTATCA	8880

55

AACCTTGTG CATAAGCCAA TTTAAAGTTT GTGGCACAAG CGAAATGTGC GTGATTCGTT 9000  
 CATTTTTAAAT CATCGTTAAA ATTTGTTCCG CATTGAATTT ATCAACAATG CGCACAGTAA 9060  
 5 AACCTTCAAT AACAGCTCTT AAAAGTACAC TGAGACCCGA AATATGATAA ATCGGCAAGA 9120  
 CAGATAGCCA ATTAGTGTCA CGATCAAATC CCAAGCTCTC TTTACATCCG ATTGCACTGG 9180  
 CATAATGATT ACGAAACGTT TGTGGCACCG CTTTTTGAGG GCCCCGTTGTC CCTGATGTAA 9240  
 10 ACATAATCGA TGCAATGTCA TCTAAATTAA ATGATGTATT TAATATGTTG GACGGCGACT 9300  
 CTTTCGGCAC CACAGTTTCA TTCGATGTTT CATATTGGAT ACCCATTGTG TTGTCCAACA 9360  
 AACTGTTTCGT TGTAATATCC CTTCCAGCGA ATTCAATATC ATCCAGCGAT ACAATTTGAA 9420  
 15 ACCCTCGTAA TTCCAGTGGC AAGGTACAAA AAATCAATTG TACATCGATT GACTTCATCT 9480  
 GATTTCGTCAT CTCATTAGGT GTCAACCTTG TATTAATCAT CGCAATTTCA ATATTTGCCA 9540  
 20 ACCAACATGC ATGTATTAAA ATGATCGATT GAATCGAATT ATCTATGTAT AGCCCAACAC 9600  
 GAGATTGTTG ATAAGCCTTG AGTCTTTTAG CCAATAGACT CGCTTCACAG TATAAATTTT 9660  
 GATAAGTATA AGATTCTTGA CCGTCTGTTA TCGCAATATG ATGTCCATTT TGTGTGCTT 9720  
 25 GTTTATATAA CCAAAAGTCC ATGCGTTATT CCTCCAAAAT CATTTACATT ATAATTATAA 9780  
 CGATTTTATG ACATTCTAGC AGTGGTTATG TTTAAAAATA TAAAAAAGTA GACGAATTGA 9840  
 TGCATTGATA TGATTGTTAT AATGCTCAAT ACATATCGTT ATATCATTCG TCTACTATTA 9900  
 30 TCAGTTATTT TTATTTAATT TTAGTGTGAT TCTGTGATTT TGATGTGGTG ATTTACCCAT 9960  
 TGTTGCCACA TCATCTGCAA TGTCAATTGG TATACGGTTC ATGTCTTGTA ATGCACTTAA 10020  
 ATGGAATACT TCATCATCTA AATTTTCAAT GAGATATACA TAATATGTTA CCTTGTCTTT 10080  
 35 TTTATATTTT AACGTTTTCC AAAAGTCCGG CTTGCAATTC AATACATTAT CCGGAATATA 10140  
 TTCAATAAAT AAGTAACGTT TGCTGCCTAC TTTGTCTATG AAATATTTTG CAGTGCCTTT 10200  
 40 TTCTATACCT CTTATATGTG CATAGTCTGC TGAAAAGTAA ATACTACCTA TTGTTTCATT 10260  
 ATGTTGTTGT ATTTCAAATC GTTGGCCTAC TATTTTATTA TTTGTGCTAC nGGGGACTTA 10320

## (2) INFORMATION FOR SEQ ID NO: 144:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1477 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

55

	GTGTGGATTG GATTTTAAAA TCACCCTCAT AAATACTGTC ATCAATATGA TAAGTTACAA	120
	TTTCACCTAT TATTAAATCA GCCCCATCTA ATACATCTCC AAGCAATATC ATTTGCGmTA	180
5	GTTTACATTG GAATCTCATT TTCGCATCTT TAATTCCTGG CGTCTTAATC GTTG TAGATG	240
	TTAAAAGTGA TAATTCTGTA CGACTCAACT CACTGTCACC ATATGCTAAC GCGCTGCAG	300
	TCTCATTAAAT ATCTTGAACA TTATCTTCGT CTGTAATATG CACAACAAAG TCTCCAGTCC	360
10	GTTCTATATT TAATGCAGTA TCTTTTCTCT TACCTCCTGC ACGTTGAACT GCAATAGCAA	420
	TCATTGGCGG ATGATTATTA ACAATATTAA AAAAGCTAAA TGGTGCTGCA TTTACTGATG	480
15	CATCTTGATT TAATGTTGTA ACAAAGCTA TAGGTCGTGG AATAATTGAA CCAATTAATA	540
	ATTTATAGTT TTCTCTAGCA GTTAATGATT GTGCATCAAA CGTATACATA ATACCTACCT	600
	CTTTTCTAAG TATATCTAGG TATTTCTCCG ATTTTGGTTA ATTTAAACAT CTATTCTCCT	660
20	CTGAAAATCA CTGTATTTA TTTAGCAAAT CTTTGTGAAAT ATGACACATA TGCATATCTT	720
	CTGGATATTT TTCTAAATGT TGCTGATGTT CTTACAGCACT TTTAATGTAG TTAGACAGCG	780
	GTAAGACTTC CACTGCAATT TGATCTCTGT CTTTACGTCG TTCAATGAAC TGACGCGCTT	840
25	CAATTAAGTG GTCATCTACA CAACTATATA AACCCGTTCTG ATACTTTTGT CCAATATCAT	900
	TTCCCTGTG ATTACACTG TAAGGATCAA TGATTCAAA TAAATAATTC ATAATGTCTG	960
	TAATTGTAA CACATGATCA TCGAAATGAA GTTTGACACA TTCAGCATAA CCATCATACG	1020
30	GACCGTCTAA TTTAGAGCTT CTTCCATTGG CTCTTCTGTC TTCTGTATGT ATAATTCCAG	1080
	GTATTGTTGC AAAAAATGCT TCAACACCCC ATAAACATCC TCCTGCTACA TAAACAACCTG	1140
35	CCATATTTAC ACCTCATCAT CCTTTTTTAT ATTTTAAACA AGGTTATACC ATTTAATACC	1200
	GCCATGACAT GATTCTGATA CACCTTCATT ACGATACCCA TATTTTTCAT AAAATGAAAT	1260
	TAATGATTCT CGACATGTTA ACGTTACACC ATGTCGATGA TGATTCTTAG CAAGAGTTTC	1320
40	AAAATAGTTT AGTAAGCGAC CTGCAATACC CTGACCTTGA TAATTGGTG CTACAACAAG	1380
	ACCTAACACA CTAATATAGC CACCTTCACT ATTATTGTG GAGACATTTT TAAATAAATC	1440
	ATCGCTAATG TAACGCTCTT TTATGACTGG ACCGTTG	1477

45 (2) INFORMATION FOR SEQ ID NO: 145:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3976 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

55

	AGGTGATTAT CCTAAAAATG CTCATGAGGT CGCTATTAAT GATAAGTTAG CTGCAGACAA	60
	CATTAGAGTC GGGGATAGAT TACATTTTAA AAATAATTCA ACTAGTTATA GAGTTTCTGG	120
5	TATTTTAAAC GACACAATGT ATGCGCATAG TTCCATTGTG CTATTGAACG ATAACGGATT	180
	TAATGCATTG AATAAGGTTA ATACGGCATT TTATCCAGTG AAAAATTTAA CACAACAACA	240
	ACGTGATGAG CTTAATAAAA TAAATGACGT TCAAGTTGTG AGTGAAAAAG ATTTAACAGG	300
10	TAATATTGCG AGTTATCAAG CAGAGCAAGC ACCGTTAAAT ATGATGATTG TTAGTTTGTT	360
	TGCTATTACA GCAATCGTTC TAAGTGCATT TTTCTATGTT ATGACGATTC AAAAAATATC	420
15	ACAAATTGGC ATTTTGAAAG CAATTGGTAT TAAGACAAGA CATTATTGA GTGCGTTAGT	480
	TTTACAAATT TTAACACTAA CAATAATTGG GGTAGGTATT GCTGTGATCA TCATAGTAGG	540
	ACTATCATTT ATGATGCCGG TAACGATGCC TTTTACTTA ACAACGCAA ATATTTTATT	600
20	AATGGTGGGG ATATTTATAT TAGTAGCGAT TTAGGTGCC TCACTATCAT TTATCAAATT	660
	ATTTAAAGTG GATCCTATCG AAGCAATTGG AGGTGCAGAA TAATGGCATT AGTCGTTGAA	720
	GATATCGTCA AAAATTTTCG AGAAGGTTTG TCTGAAACAA AAGTTTTAAA AGGTATTAAT	780
25	TTTGAAGTGG AACAAAGGGA ATTTGTCATT TTAAATGGTG CCTCTGGTTC TGGGAAAACA	840
	ACATTGCTAA CGATATTAGG CGGATTGTTA AGTCAAACGA GTGGTACAGT GCTTTACAAT	900
	GATGCGCCAT TGTTTGATAA ACAGCATCGT CCTAGTGATT TACGATTGGA AGATATTGGT	960
30	TTTATTTTTC AATCTTCACA TTTAGTTTCT TATTTAAAAG TGATAGAGCA ATTGACACTC	1020
	GTAGGTCAAG AAGCGGGAAT GACCAAACAA CAAAGTTCAA CAAGAGCAAT ACAACTTTTG	1080
35	AAAAATATTG GTTTAGAAGA TCGCTTGAAT GTATATCCGC ATCAGTTATC TGGCGGTGAA	1140
	AAGCAACGTG TTGCGATTAT GAGAGCATTT ATGAATAATC CGAAAATCAT TTTAGCAGAT	1200
	GAGCCACAG CAAGTTTAGA TGCCGATAGA GCAACAAAAG TTGTTGAGAT GATACGTCAA	1260
40	CAAATTAAAG AACACAAAT GATTGGTATT ATGATTACAC ACGATCGAAG ATTATTTGAA	1320
	TATGCAGATC GAGTGATTGA ATTAGAAGAT GGCAAATAA CTGATTAGTG GCTTGTAAG	1380
	ACGCTAAATG TTAATGATTT AAGACATAGT AGTATAAAG TTAGATAACA GAATACGATT	1440
45	TGGGTTTACA AAAACAGGC TGGGACATTA AGTTCCTTAGG CAATGTAAAA AAGCTGATTT	1500
	CTATTAATTA TTTGATAGAA ATCAGCTTTT TTGATATGTA TTTTATAATG TACAGCTCGT	1560
	TGCATTCATA TAGCTTGAAG TCACGTTTAA AACCATATCT ATCATTATGG TATGCATATC	1620
50	TTTTAAACC TATTCTTTTG TTATTAGGAC ATATAAATTC ATCATTAGT TCGTCATATT	1680
	TCCAATTTTG AGTGTTAAAA ATGTCACTTT TAACTTTCT AGTTTTATCT TTAATAAACA	1740
55		

EP 0 786 519 A2

	CACTATCATA ACATGCATCA GCTACAATAT ACTCCGGTAA ATAACCGAAG nTATTTTgAA	1860
	TCATTGTTAA AAATGGAATT AAAGTTCTAG TATCTGTTGG GTTTTGAAAT AGGTCATAGG	1920
5	ATAAAACAAA TTGAGAATTT GTCGCTATTT GTAAATTGTA TCCTGGCTTA AGTTGGCCAA	1980
	AGTGTCTTAT TTTTTTAAAG TATTTAAAAG TAAAATTACA TGTTAATACG TAGTATTAAT	2040
	GGCGAGACTC CTGAGGGAGC AGTGCCAGTC GAAGaCAGGG GCCCCAACAC AGAArcTGAC	2100
10	ATATAGTCAG CTTACAACAA TGTGCCGGTT GGGGTGGCTG AGACGGCACC CTAGGAAGGG	2160
	ACCCGTCATC AAAAATTCTA TTTATAGAAT TTTACAGTAA TGTGCCAGAT GGGCATAGCG	2220
	AAgcCATTCA ATACGAAGTA TTGTATAAAT AGAGAACAGC AGTAAGATAT TTTCTAATTG	2280
15	AAAATTATTT TACTGCTGTT TTTTTTAGGG ATTAATGTCC CAGACTCTTT AGTTTATTTA	2340
	TTTTCAATAT AACAATTGTC TAATCAAGGA TTAACGAATA TTTAAAGATA GTTTGACGCA	2400
20	ATATTAGAAA CAACCTATAA TAATAGTTTG TTTGTGGATT AACTATTATA AATAAAAGCG	2460
	GCGTAAAGAC ATATAAACCA ACTACTTGAA CAATATAACG TTAATAACAA TCTATACTGA	2520
	TACATTACGC CTAGATAATC TTTGATGAGC ACATGTAAGA AAAAGTGATA TGGTGTATGA	2580
25	CTTCGACAC CATCGATAGA TAAACCTAAT TTTTGGGCTA GTCGTAAGGC GCGCAATACA	2640
	TGAAACTGAC TTGTtACACA AACAATTTTA ACTGCTTCAT GATACAAATT GTTGATGATT	2700
	TGTTTAGAAT ATAAAAAGTT TGTGTATGTA TTTATAGAGT GAGATTCCAT TAGTATATCT	2760
30	GTTTTATCAA CACCATGTGC AATCAAATAA CGTTGCATAG CTAAAGCTTC AGAAATTGGT	2820
	TCGTCTGGTC CTGTCCGCC AGATACAATG ATCTTTGTTG CTGATGCTTG TTGTTGATAG	2880
	ATATCAAGTG CACGATCTAA ACGCGCTGCA AGCATTGGTG TGACAAATTC GGTAAAAATA	2940
35	CCAGCACCTA ACACAATTAT GATATCAACT TCTTTGTTGT ATGATCTATG TCTATATGAT	3000
	ACTGtCCAAA CGAGATAACA AATAAAGGTT AGTAACAGGG AAAGACATAA TATAGCTAAC	3060
40	CACATAGACA AACCTTTCAC AATAGGTGAC TGAATCGTAC TTATAAATAG AAGTGCTGAT	3120
	GTGTAGAGTA CAAATTTATA TGAAAAAGAT AATAATTTTT TAATAAATAA GCGACTAGAA	3180
	GTATGAGAAA ATAAATATCT ATGTTTGAAT AGCATGATAA TACTGATTAT TATAAATGTT	3240
45	ACAAACATAG ACCAAGGGAA AGTATAGGTC ATGATGCTAT AGATGAGTGA CAAAAATATC	3300
	GATATGACAA CTAAGATGTA GCATGTTAAA TTTAACGTCA GAGTATAGTT GAAAATTAAC	3360
	GGACAAATAA CGATAAGTAT AAATATTAAT AATAAATTCa ATAACATACT GACACCTCGC	3420
50	TTATAATAAA TATTAAATAT AAATGTAGAT GATTTAATTT ATTAAAGCAA GGAGAAAGCA	3480
	GCAACATGTA AATCTTAATT TGTTATATTA TATATGGGTC AATATTTTTG TGTTTTTTAG	3540

55



TATGGTAAAA CATTTACAAG ACCATATTCA ATTTTITAGAG CAGTTTATAA ATAACGTAA 3660  
 CGCATTAAGT GCAAAAATGT TGAAAGATTT ACAAATGAA TATGAAATTT CATTAGAGCA 3720  
 5 GTCTAACGTA TTAGGTATGT TAAATAAAGA ACCTTTGACA ATTAGTGAAA TCACGCAAAG 3780  
 ACAAGGTGTA AATAAGGCCG CAGTAAGCCG ACGAATTAAA AAGTTAATCG ATGCTTAATT 3840  
 AGTTAAGTTA GATAAACCAA ATTAAATAT TGATCAACGT TTGAAATTCA TAACCTTAAC 3900  
 10 TGACAAAGGT AgAGCATATT TGAAAGAACG TAATGCGATT ATGACAGATA TTGCGCAAGA 3960  
 TATTACTAAT GATTTA 3976

## (2) INFORMATION FOR SEQ ID NO: 146:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3346 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

25 GCTACCTAGG CATTTAAGAG ATCAAAAAAT GTATGAATAT GAACGTTATT TTTATGAGCA 60  
 AGAACTTAAT GCGCTTGATG aAGGGGAAAT TTTAAAGAAG TTTAAAGACC CACAAGATGT 120  
 TGCAGCTGAA ACAAAGCTA GAAGTGTTAT TGATTATGCT GAATCTAAAC CAACATTTGA 180  
 30 AAATATTTCA AGAGCTGTTG CTGCTTCATT AAGTTTAGGC ATTCTATCTA TTTTGTGTCAT 240  
 CCTTATACCA GTATCTATAG TTGGATTATT TGTATTAGCA TTATTTTAA TATCACTTTT 300  
 GCTGCTGTTT TGTCCAATTA TTTTATTAGC ATCAGCAATA TCCAGAGGAA TTGTGGACTC 360  
 35 AATTAGTAAT GTATTTTTTG CCATATCATA TTCAGGATTA GGATTAGTAT TTATCATTGT 420  
 CATATTTAAG ATTTTAGAAT ACATTTATCG TTTAATCTTA AAATATTTAC TTTGGTATAT 480  
 TAAAACTGTC AAAGGAAGCG TTAGAAAATG AAGAAATTCT TTTTATTGG GCTTTTAGTG 540  
 40 TTTGTTGTCT TTTTACAGC AGCAACCATT ATTTGGTTCA GCTATGATAA AAACAAATAT 600  
 GGTACTAAAC AATATGATAA AACATTCAAA gACGATGCTT TTGACAATGT ATCTATAAAT 660  
 45 TTGGATAGTA CAGAACTTCG TATAAACCGG GGAATCAAT TTAGAGTTAA ATATGATGGT 720  
 GACAATGATA TATTAATTAA TATAGTAGAT AAGACGTTGA AGATTAGTGA TAAAAGGTCT 780  
 AAGACAAGAG GATATGCAAT TGATATGAAT CCTTTTCATG AGAATAAGAA AACGTTAACG 840  
 50 ATTGAAATGC CTGATAAAAT GATTAAACGT TTAAATCTAT CATCTGGAGC AGGAAGTGTT 900  
 AGAATCAGTG ATGTTGATTT AGAGAACACA AGTATTCAAA GCATTAACGG TGAAGTAGTT 960

55

	AGTAAAAGTA ACATTAAAAA TAGCAATATT AAAGTTGTTA TTGGTACGCT ACAAATCGAC	1080
	AAGAGTCAAA TTAAACAATC CATATTTTAA AACGATCATG GTGACATTGA ATTTAAAAAC	1140
5	ATGCCATCAA AAGTAGATGC AAAAGCTTCT ACTAAACAAG GAGATATTCG TTTTAAGTAT	1200
	GATAGTAAAC CTGAAGACAC TATACTAAAG CTAAATCCGG GAACGGGTGA TAGCGTAGTT	1260
	AAAAATAAAA CATTACTAA TGGTAAAGTT GGGAAAAGCG ACAATGTTTT AGAATTTTAT	1320
10	ACGATTGATG GTAATATCAA AGTTGAATAA ATAAAGGATG TAAGCACCGA TATTAGGAAG	1380
	CATAATTTCT CTAATATCGG TGTTATTTAT TTGTTGGCAA AAGTTAAGTC GGTATCTATA	1440
15	TTGCCAGTAA AGTGAGTGAT ATTAAGGTCT TGACCATCTA ACCATGATTT GAAATCTATT	1500
	ATTTCTGGTG GCGCATTTTC TCCCAATGTA AAATATGCAG TTAATGTTTC AGGTTGATAC	1560
	ATTGATGTAT GGATGGTGCC AGACCAGCTT TTGAATAGTT TACTGTAAAT TTCATACTGA	1620
20	GGATTATTGA ATAACTTAAA TGCTGTAGTC ATATCTAAAT TATCATTAGT TTGTGAAATG	1680
	GTACGCGCCA GTCTTTCTTT AGATTCTTTT GTATAATTAC GATTTTCATG TGTTAATATT	1740
	TCAAATGAT TTGTACATAT ATTATCATAA CGAACATCTA TTGATCTCGG TGCTACTTCA	1800
25	ACAATTGCAT GGTTCATGA TTTGTCCATC AGTATGTAGC TAAATGAGCT TCTGTGTGGT	1860
	ATTTCTTTCA ATAATTGGAT TGCTTCTGTT ACATTTCCGC AATTTTCAAG AATTAGACGA	1920
	CCAATCATAT AACATACAAA ACCATTGCT GGTTCCTCC GGTGCATAAA GTTATAGCCC	1980
30	ATAGTTAATC CTGACTCATT CATACCATCC ATTCTTCCAG TTACCCTTGA TACAGGACCA	2040
	ATTTGAGCTA AACCGCTATC TGTAGGTTGA TAAAGTAAGT AGCGACCATC ATAAGTTGCA	2100
	GGGTGGTAAT CATAATTTCT AACCATGAAG TCTTTGCCTT GAAAGACCGT GCAaCCACTT	2160
35	TCTTTTAAAT CGGTAAAACG ATAATGTCCA AAGTTTAAAA TAATTTGGCG TGTTGGCATT	2220
	TTGAGTATAC TTTGTAGTCC CATTAATTCT TCCCATATTT GAGGTGCGTA TGTTTGAAT	2280
40	ATTTGATAAG TTTCAATTAC ATCTATATCG AAACGTGGGA CaCnTTTTTT CCATTCTTTT	2340
	TCTCGATTTT TTAGAAGAGG TGTTTGTGTA AGCCATTTAC CAGTTTTAAC ACCTAACTCG	2400
	AAATGTGAAC CTCTAAAAGT CATGATATCT GATGTCACCT GTTGCATATC ATCGGCCCTT	2460
45	TTCTTTTTAG TTGTAATATA TTGTAAATAA ATAGTAATCG TATGTATATT GAATGTCATG	2520
	TTAAATAAAG TTATATTTTA CTAAATGAAA TATAAAATTG TTTGAGGTGA TTTCTCGGTG	2580
	TATAAGACTT ATCAATCAGT TAAACATAT TTTTATAGAT GGTGGGGATA TTGAGTTAAA	2640
50	AACTTAAAT CATCTTATCA TAAATATCAA TCTTAAGTTA GCATTACGA TAATAGTCAT	2700
	TGTTAACATT AGCATATAAG GTCATGTCAC GTTGAAACAG AGGTTCCCTCG GCATTTTTGA	2760

55

	TTATTTAATG ATTATTCTAT ATATGATAGT ATAATGAAAT GTAGATAGGT ATTTAATTTA	2880
	ACAGAGGTGA AATTGAGATG TGGAATTTTA TTAAATGtGT GkTTAAATTC GTATTTAGCT	2940
5	TAGTTGCTAT TACAACATTA GTTGCTGGTG TTGGTGTAGT AGCATTTGCT TATATCTTTA	3000
	AAAAAGATTT TGAAGATATT GAAAGAAAAA CTAAAGAAAT TATTTCTGAT ATTGAAAGTA	3060
	AAAATAACTA ATAACATTTA GAGGCTGGGA CATAAATCCC TAAAAACAG CAGTAAGATA	3120
10	ATTTTCAATT AGAAAATATC TTAAGTCTGT TCTCTATTTn ATcAmTACTt CGTATTGAAT	3180
	GGCTTCGCTT TCCTAGGGTG CCGTCTCAGC CTGGTCTTC GACTGGCACT GCTCCCTCAG	3240
	GAGTCTCGCC ATTAATACTA CGTATTAAACA TGTAATTTTA CTTTGGAAT ACTTTTAAAA	3300
15	AATAAGACAC TTTGGCCCAA CTTGGCACAT AAATGTAAAA TTCAAT	3346

(2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2375 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

	GTTGAAGAAA GAAATATAAC AGTCAATTAT AATTATAACC TTGTTGAAAT CGACGGTGAC	60
30	AAAAAGTGG CTACATTGCA ACATATCAAA GCATACGATA GAAAAACAAT AAGTTATGAT	120
	ATGTTACATG TAACACCACC TATGGGTCCC TTAGATGTAG TAAAGAAAG TACACTTTCA	180
	GATAGTGAGG GTTGGGTAGA TGTTAACCCA ACCACATTAC AGCATAAAAG CTACTCTAAT	240
35	GTATTTGCAC TTGGTGATGC TTCAAATGTA CCTACTTCAA AAACAGGCGC ACTATTcGTA	300
	AGCAAGCACC TATCGTCGCT AATAATTTAT TGCAAGTGAT GAATAATCAA ATGTTAACGC	360
40	ATCATTATGA TGGTTATACT TCATGCCCTA TTGTTACTGG ATATAATAGG TTAATACTTG	420
	CAGAGTTTGA TTATAATAAA AATACTAAAG AAACAATGCC GTTTAATCAG GCCAAAGAAC	480
	GTAAGAAGTAT GTATATATTT AAGAAAGATT TATTACCTAA AATGTATTGG TACGGCATGC	540
45	TAAAGGATT AATATAATAA AGTACAGAAA ACAATAAATT TTTAATGAAA AATCTTTTAC	600
	TATAAAGAT TAAGTATTTA AATGACGTGT CAGTGTGTG TTTATATGTC GTGAATTTTT	660
	AGCTCTAAAT AGTATAAGAT TGAAAAAGTT GTTACTGTTT TAAATGATCA CGATGAAGTC	720
50	ATTCAATAAG AATGATTATG AAAATAGAAA CAGCAGTAAG ATATTTTCTA ATTGAAAATC	780
	ATCTCACTGC TGTTTTTTAA AGGTTTATAC CTCATCCTCT AAATTATTTA AAAATAATTA	840

55

	AGATATTCAA ACCACGTGTA CTCAAAATGA TAGCTTGGTA TGTACCTCCA ATAGTAATTT	960
	CAATAACTTT GTCTGTTGAA CACTAAGAGC AATTTTAATT TCATAATGTG TTGTAAACAT	1020
5	TTTTTTTGAT TGGAGTTTTT TTCTGAGTTA AACGATATCC TGATGTATTT TTAATTTTGC	1080
	ACCATTTCCA AAAGGATAAG TGACATAAGT AAAAAGGCAT CATCGGGAGT TATCCTATCA	1140
	GGAAAACCAA GATAATACCT AAGTAGAAAG TGTTCATCC GTGTAAATT GGGAAATATC	1200
10	ATCCATAAAC TTTATTACTC ATACTATAAT TCAATTTTAA CGTCTTCGTC CATTTGGGCT	1260
	TCAAATTCAT CGAGTAGTGC TCGTGCTTCT GCAATTGATT GTGTGTTTCA CAATTGATGT	1320
15	CGAAGTTCGC TAGCGCCTCT TATGCCACGC ACATAGATTT TAAAGAATCT ACGCAATCTC	1380
	TTGAATTGTC GTATTTTCATC TTTTCATAT TTGTTAAACA ATGATATG CAATCTCAAT	1440
	ATATCTAATA GTTCCTGCT TGTGTGTTCT CGTGGTTCTT TTTCAAAGT GAATGGATTG	1500
20	TGGAAAATGC CTCTACCAAT CATGATGCCA TCAATACCAT ATTTTCTGC AAGTTCAAGT	1560
	CCTGTTTTTC TATCGGGAAT ATCATCGTTA ATTGTTAACA ATGTGTTTGG TGCAATTCG	1620
	TCACGTAAAT TTTTAATAGC TTCGATTAAT TCCCAATGTG CATCTACTTT ACTCATGCGT	1680
25	TTGATAAAAA CTAAATAAT ATTAATTCGG TCATCAGTGG CGTTAAATCT TTTATCATTT	1740
	TTAGTTATAG TTGATAAATT TATATTTATA AGCATATATG GATATTTTCA CAAAAATTTT	1800
	TATTTATATA AATCCGAAC GCATACATAT TTGTTTAAAT AAGAGGTATT ATTTTTCGGG	1860
30	AAATTGCTGT CTGAGTTAAA AGGATTAGTT TTATAAAATG AGTTGAAC TAAGCCAAAA	1920
	CGATTAAAAT ACTGATAATC CATTTTGTGA TTATGTTAGG GACTTTTTTA CTTAATTTTA	1980
35	ACCCTATTGG aGcMAATATA ATACTCCCTA TTATAAGGAA TAAGGCGTCA TATAAGGGA	2040
	TATAACCTTG AATAAGTTTG ATGACAAAAG CACCAATTGA AGATATAAAA GCAATTACTA	2100
	TACTATTAGC GACTACAGTA TTCATTGGTA ATTTGAATAA AACCAATAAT ATAGGAATAA	2160
40	TAATGAAGGC ACCACCTGCA CCTACTATAC CTGAAATAAT ACCAATGAAA AGGCCAATGA	2220
	TAAC TAATAA ATATTTATTA AATGAAGACT TTTCGGAAC AGGTTTCACT TTAATAAACA	2280
	TTAATGTTAA TGCAAGTAAA GCAATAATGA TATATACCGT ATTTACAAAT GTAGCATCAA	2340
45	ATAAATTTGC TAGAAATGCA CCTAACATAC TCCCT	2375

(2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6115 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

	GAGGTTTCTA GACAAGCTTT TAATAACTTA CCAAACATCAT TAAGrTGGTT gTgTGGACT	60
5	GCCTATTATC mAAGtATTAT GaGTTGTTTA ATATTAGtGC TAArACATAC GAAGAGTGGT	120
	TTAAACAATT TAGTAGTAAG AAAGCACAAT TCAGTATTAA TCTCACGGAT AAATGGATAA	180
	TTCAAATCGC ATATGGTAAA TTAATAATAA TGGCTAAAAA TAATGGCGAT ACATATTTTA	240
10	GAGTTCAAAC AATTAAAAAG CCAGGTAATT ATATTTTTTA CAAATATCGA TTAGAGATAC	300
	ATTCTAATTT ACCAAAATGT TTATTTCCGC TTACAGTGAG AACACGACAA AGTGGCGATA	360
	CATTTAAACT GAATGGGCGC GATGGTTATA AGAAAGTGAA TCGCCTGTTT ATAGATTGTA	420
15	AAGTGCCACA GTGGGTTCGG GATCAAATGC CAATCGTATT GGATAAACAA CAGCGCATT	480
	TTGCGGTAGG AGATTTATAT CAACAACAAA CAATAAAAAA ATGGATTATA ATTAGTAAAA	540
20	ATGGAGATGA ATAGCGTTAT GCATAATGAT TTGAAAGAAG TATTGTTAAC TGAAGAAGAT	600
	ATTCAAAATA TCTGTAAGGA ATTGGGAGCA CAATTAACAA AGGATTATCA AGGTAAACCA	660
	TTAGTATGCG TGGGTATCTT AAAAGGCTCA GCAATGTTTA TGTGAGATTT AATTAAACGA	720
25	ATTGATACCC ATTTATCAAT TGATTTATG GATGTTTCTA GTTATCACGG AGGCACTGAG	780
	TCAACTGGTG AAGTTCAAAT CATTAAAGAT TTAGGTTCTT CTATTGAAAA TAAAGACGTA	840
	TTAATTATTG AAGATATCTT AGAGACTGGT ACTACACTTA AGTCAATTAC TGAATTATTA	900
30	CAATCTAGAA AAGTTAATTC ATTAGAAATA GTTACTTTAT TAGATAAACC AAACCGTCGT	960
	AAAGCGGACA TTGAAGCTAA GTATGTAGGT AAAAAAATAC CAGATGaATT TGTGTGGT	1020
	TACGGTTTAG ATTATCGTGA ATTATACCGA AACTTACCAT ATATCGGTAC GTTAAACCT	1080
35	GAAGTGTATT CAAATTAATT TTTAATCAA TTTCAGTTAT TATTACTATG CGTTTGAGAA	1140
	ATAATAGTGT AGACTCAAAA ATATGAAAAA TGTATTTTAT ATATATTAA TTTTAGACAA	1200
40	GACATATGTC TTGAAAAGTT GAAAAATATA GAGATTGATA AAACATAAC GGGTGTGAAT	1260
	GACATTGATG TTAAGCTCAA TTACTAGCTT ATAAACATG TCATATGTTA CAATTTTGT	1320
	TAGTTTTATT ATGGGAAGTA GGAGGAAATG ACGCATGCAG AAAGCTTTTC GCAATGTGCT	1380
45	AGTTATCGTA ATAATAGGCG TTATTATTTT TGGTCTATTT TCATATTTAA ACGGTAATGG	1440
	AAATATGCCG AAACAGCTTA CATATAATCA ATTTACTGAG AAGTTGGAAA AAGGTGACCT	1500
	TAAACTTTA GAAATCCAAC CACAACAAA TGTCTATATG GTAAGTGGTA AAACGAAAAA	1560
50	TGATGAAGAC TATTCATCAA CTATTTTATA TAACAACGAA AAAGAATTAC AAAAAATTAC	1620
	TGATGCTGCT AAAAAGCAAA ACGGTGTAAT ATTAACGATT AAAGAAGAAG AAAAACAAAG	1680
55		

## EP 0 786 519 A2

TTTCTTCCTA AGCCAAGCAC AAGGTGGCGG TAGTGGCGGT CGTATGATGA ACTTTGGTAA 1800  
 ATCTAAAGCA AAAATGTACG ATAATAATAA ACGTCGTGTT CGTTTCTCTG ATGTAGCAGG 1860  
 5 GGCAGATGAA GAAAAACAAG AATTAATTGA AATTGTTGAT TTCTTGAAAG ATAATAAAAA 1920  
 ATTCAAAGAA ATGGGATCTA GGATTCCTAA AGGTGTCCTA CTTGTTGGAC CTCCAGGTAC 1980  
 TGGTAAACA TTACTTGCTA GAGCGGTTGC AGGTGAAGCT GGCGCACCAT TCTTCTCTAT 2040  
 10 TAGTGGTTCA GACTTTGTAG AGATGTTTGT TGGTGTGGT GCGAGCCGTG TTCGTGACTT 2100  
 ATTCGATAAT GCTAAGAAAA ACGCGCCTTG TATCATCTTT ATCGATGAGA TTGATGCTGT 2160  
 TGGTCGTCAA CGTGGTGCAG GTGTTGGTGG CGGTCATGAT GAACGTGAAC AAACCCTAAA 2220  
 15 CCAATTATTA GTTGAAATGG ATGGTTTCGG TGAAAATGAA GGTATCATTG TGATAGCTGC 2280  
 TACAAACCGT CCTGATATCC TTGACCCAGC CTTATTACGT CCAGGTCGTT TTGATAGACA 2340  
 AATTCAAGTT GGTCGTCCAG ATGTGAAAGG CCGTGAAGCA ATTCTTCATG TTCATGCTAA 2400  
 20 AAACAAACCA CTTGATGAAA CGGTGATTT AAAAGCAATT TCACAACGTA CACCTGGTTT 2460  
 CTCAGGTGCT GATTTAGAGA ACTTATTAAT TGAAGCATCT TTAATTGCTG TACGTGAAGG 2520  
 25 TAAAAAGAAA ATTGACATGA GAGATATCGA AGAGGCAACG GATAGAGTTA TAGCCGGACC 2580  
 TGCTAAGAAA TCTCGAGTTA TTTCTAAGAA AGAACGTAAT ATTGTTGCTC ATCACGAAGC 2640  
 TGGTCATACA ATTATCGGTA TGGTACTTGA TGAGGCAGAA GTAGTGATA AAGTTACTAT 2700  
 30 TGTTCCACGT GGACAAGCAG GTGGTTATGC AATGATGCTA CCTAAACAAG ATCGTTTCTT 2760  
 AATGACTGAA CAAGAGTTAT TAGATAAAAT CTGTGGTTTA CTTGGTGGAC GTGTATCAGA 2820  
 AGATATTAAAC TTTAACGAAG TATCAACAGG TGCTTCAAAT GACTTCGAAC GTGCAACACA 2880  
 35 AATCGCACGC TCAATGGTTA CGCAATATGG TATGAGTAAA AAATTAGGAC CATTACAGTT 2940  
 CGGTCATAGC AATGGTCAAG TATCTTAGG TAAAGATATG CAAGGTGAGC CTAATTATTC 3000  
 AAGCCAAATC GCATATGAAA TTGATAAAGA AGTTCAACGA ATCGTTAAAG AACAAATACGA 3060  
 40 ACGTTGTAAA CAAATTTTAT TAGAGCACAA AGAACAAATTA ATTTTAATTG CTGAAACATT 3120  
 ATTAACAGAA GAAACATTAG TTGCTGAACA AATTCAATCA TTATTCTACG AAGGTAAATT 3180  
 ACCTGAAATT GATTATGATG CAGCTAAAGT TGTTAAAGAT GAAGATTCTG AATTTAATGA 3240  
 45 TGGTAAATTC GGTAAATCTT ATGAAGAGAT TCGTAAAGAG CAATTAGAAG ATGGACAACG 3300  
 TGACGAAAGT GAAGATCGTA AAGAAGAAAA AGATATTGCT GAGGATAAAA AAGAAGCTGA 3360  
 50 TAAATCTGAT GAAAAAGATG AACCAGCACA TCGACAAGCC CCAAATATCG AAAAACCTTA 3420  
 CGATCCAAAT CACCCAGACA ATAAATAATC GATTATATTC AGTACCTCTT TCTATGATAA 3480

55

EP 0 786 519 A2

	AATTGTTATA GCAGAAAATA ATTGTAAAAC AAGTTACTTC ATTATTTAGA ATGATGGGTG	3600
	TAGAATAAGT ACAATTGTTG CATTTTATGA AGTAAAGTAA TTTTAAAT ATAGAGTAAT	3660
5	AGAGGAGATT GAAATAATGA CACACGATTA TATTGTTAAA GCATTAGCAT TTGATGGAGA	3720
	GATTAGGGCT TATGCTGCTT TGACAACCTGA AACTGTTCAA GAAGCACAAA CGAGACATTA	3780
	TACATGGCCG ACAGCATCTG CTGCAATGGG AAGAACAATG caCAGCAACA GCTATGATGG	3840
10	GCGCAATGTT GAAAGGTGAT CAAAAATTAA CTGTCACCTGT AGATGGCCAA GGACCTATTG	3900
	GACGAATTAT TGCCGATGCA AATGCTAAAG GCGAGGTGCG TGCTTATGTA GACCATCCAC	3960
	AAACTCATT TCCATTAAAT GAGCAAGGTA AACTTGATGT AAGACGAGCG GTAGGGACAA	4020
15	ATGGATCTAT TATGGTTGTT AAAGACGTTG GAATGAAAGA CTATTTCTCT GGAGCAAGTC	4080
	CaATTGTTTC AGGAGAACTT GGTGAAGATT TTAATTATTA TTATGCTACA AGTGAACAAA	4140
	CACCTTCATC GGTAGGTCTT GGTGTATTGG TAAATCCTGA TAATACGATT AAAGCAGCAG	4200
20	GAGGATTTAT CATTCAAGTT ATGCCAGGTG CCAAAGATGA AACAATTTCA AAATTAGAAA	4260
	AAGCAATTAG TGAAATGACA CCAGTTTCTA AATTAATTGA ACAAGGATTA ACGCCAGAAG	4320
25	GATTACTAAA CGAAATCTTA GGTGAAGACC ATGTGCAAAT TTTAGAGAAA ATGCCTGTTT	4380
	AATTTGAATG TAATTGTAGT CATGAGAAAT TTTTAAATGC TATTAAAGGA TTGGGCGAGG	4440
	CTGAGATTCA AAATATGATT AAAGAAGATC ATGGTGCTGA AGCAGTATGT CATTTCTGTG	4500
30	GAAATAAATA TAAATATACT GAAGAAGAAT TAAACGTGTT GCTAGAAAGT TTAGCGTAAT	4560
	TTAATTTAAA TCAATACGCT AAAATGTTTA TTTTAGCGG TTTAGTGAAA TGTAGAACTA	4620
	AATAGTTGTA TAATCCTTAG TGATTTTGTT TGCTTTCTAG AATTTATTTG ATAAAATAAT	4680
35	TCTATATCCG ATAAATAAAC TAAGATTTCA ACAACTAACT AAAAAGGAGT GTTCTTAATG	4740
	GCAAAAAAC CAGTAGATAA TATTACTCAA ATTATTGGCG GTACACCGGT AGTCAAATTG	4800
40	AGAAATGTAG TAGATGACAA TGCAGCAGAT GTTTATGTAA AATTGGAATA TCAAAATCCA	4860
	GGTGGTTCTG TAAAGGATAG AATTGCTTTA GCAATGATTG AAAAAGCAGA GCGAGAAGGC	4920
	AAAATTAAAC CTGGCGATAC AATTGTAGAA CCAACAAGTG GTAATACAGG TATCGGTTTA	4980
45	GCATTTGTAT GTGCTGCTAA AGGATATAAA GCAGTATTTA CTATGCCCGA AACAATGAGC	5040
	CAAGAGCGTC GTAATTTATT AAAAGCATAC GGTGCGGAAT TAGTTTTAAC GCCTGGATCA	5100
	GAAGCGATGA AAGGTGCAAT TAAAAAGCT AAAGAATTGA AAGAAGAACA TGGTTACTTC	5160
50	GAGCCACAAC AATTTGAAAA CCCTGCGAAC CCTGAAGTTC ATGAGTTAAC TACAGGTCCT	5220
	GAGTTATTAC AACAATTGA AGGGAAAACT ATCGATGCGT TCCTAGCTGG TGTGGTACT	5280

55

GTTGCTATAG AGCCTGAGGC TTCTCCAGTA TTGAGCGGTG GTGAGCCAGG TCCACATAAA 5400  
 TTACAAGGTT TAGGTGCTGG ATTTATTCCA GGCACTTTGA ATACAGAAAT CTATGACAGT 5460  
 5 ATTATTAAG TAGGAAATGA TACAGCGATG GAAATGTCTC GTCGAGTTGC TAAAGAGGAA 5520  
 GGTATTTTAG CAGGTATTTT ATCAGGTGCT GCGATTTATG CTGCCATTCA AAAAGCAAAA 5580  
 GAATTAGGAA AAGGTAAAAC AGTAGTAACA GTATTGCCGA GTAATGGTGA ACGCTACTTA 5640  
 10 TCAACACCTT TATATTCATT CGATGACTAA TTAATGTCAT TTAAGAGAGT GAGTTATCTT 5700  
 TTTGAGATAA CTGCTCTTT TTTTCTACCA TGTATATTTT TAAAAATATG AGCGTTAAAT 5760  
 TAAACATTTT TCTGATAAAA ATATCCAGTG AATGATAAGA TAATAAACGT ACATACTAAT 5820  
 15 AACTAGTAAA TAGCAGGAGT AAATTTTATT AGAGTTAAAC AATACATAAT TAAAGGGTGG 5880  
 TTAACATGAC TAAACAAAA ATTATGGGcA TATTAAACGT CACACCTGAT TcATTCTcAG 5940  
 20 ATGGTGGAAG ATTTAATAAT GTTGAATCAG CTATAAATAG aGTGAAAGCC ATGATAGATG 6000  
 AAGGTGCTGA CATTATAGAT GTTGGAGGTG TTTCAACGAG ACCCGGTCAT GAAATGGTTT 6060  
 CATTAGAAGA TGAGATGAAC AGAGTATTAC CTGTTGTTGA AGCTATTGTC GGTTT 6115  
 25 (2) INFORMATION FOR SEQ ID NO: 149:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10401 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:  
 35 TAGATACTGG GnTAAAcATc AAAAATAtyT GcTtATTCaC GTGTTTAcGc TCCcTCAAAC 60  
 GCAACGTTAA TTGCGTGTA TcATTTAGTG TGAATcAGA CGCTTCITCC ATGACTATGT 120  
 40 CTGATATGCC TTTTATCGAC TTTATTTTCT CTGGGTTATC TAATCCTTTA AACAAAAAAA 180  
 CTGCGCCGTT TGGCAATTCA ACTTTGTTAT CAGTCTTATT CCAAAGGCAC ATGTCCCAA 240  
 TACCAAAGTT TATCAAACAA TCTTTAACAT CTTCGAACAA ACTATCTTTA ATTGTTGATT 300  
 45 GTACTTTTCT AAGCCACAGT ATACGCCTAG GATATTTCCA ATCTTGCAAT GCTTTGAGTA 360  
 CAACTTTTGG TATAACGCCG TGAGACTTAC CGCTCGAACC TCCACCGTAA TGkACTTCAG 420  
 TGAAGTcATC GTAATTGGTT AGTATTTTGA ATATGTTTCT ATTGAAAACA TTAGACGGTT 480  
 50 TGTAAAGTT TAATTTAACT TTCGTCATCG TACTCACCAA TATTAATCTC AATATTCTTC 540  
 TGAGTAATTT CTTTTTTATC GATATACGCA CCATGTACTT TTAGTATGTG GTCAATAGAT 600

55



	TTTAAATGGT CATATTTCTT ACTGTAAGCC TCTTGAGGTT CTCCTCTAGC AATAGAAGCA	720
	GATAACGCTA AAGCTTCTGT AATACTCATT AAACGCTCTT CTTGTATCTG TTCTAATCGT	780
5	TCTTTAATAT ATTCCGAAAC ATTAACATTT CTTAACAATC GACTTGCTAA AGACTCTGCT	840
	GTTTTCTTAC TATAACCTGC TGTAATTGCT GCTTTTTTAC CATTACATCC ATTCATTATA	900
	TATTCATCTG CGAATCTCTT TTGTTTTTCG TTCATTTTCAT TTACCACCAA CTCTCGCGCT	960
10	ATACGCTTTT TAAATTTAAA AAAGGATTGG CTATAATCAG CCAACCCACA TAGATCCTTT	1020
	ATTCCTAATT GCGATAAGGG AAACGCAGTA CGATAGTCAA TATCCTACAC TATCATAATA	1080
	TCTCATTTAA GGTATCAAAA ACTGCCACTT TACTGCCAAT TTCAGTCTTC CCCTAACTCT	1140
15	TCCGCCAATC TAGATATGAT TTTTCTTTTG ATTCTATGAG CAGTTCTATC AGAAATGTGT	1200
	ATGTCAACAC AAACTTTCAC TAATTCCTTT TTATTAAAAT AATACTCTTG AATGAATTCTG	1260
20	CGTTCTTTCC TGCTTGATGT GTTGATTATA CGTTCAATAG CGCTCTTAAA CTCAAGGATT	1320
	TTACCTCTTC GTATACTACA AAGATAATTA GTTACTGCCA TTTCTGTTTT CGATGTATTA	1380
	GACGGTACAA ACTCCCCGCC TATATTTGTA TCTGTTGGAA TCCACGGTGT CATTATTTCA	1440
25	CTTCTTAAAT CTTCAAGTTG TTTATGATAA TTAGGATAAT CACACAATC ATCTTCTAAC	1500
	TTTCGAACTG TTGATAATTT TAATCCGTAT TTCTTTTTAG TCATGAATAC CCTCCGTACA	1560
	AATATGTTTA ATCTTCAAAG TGTCTCAATC TACTTCTTAA TATCTCTATC TCTCGCTCTT	1620
30	TAACTTTAC ATCACCTTTT AACTGTTCCG CTTGTAACAT CACACCAAAC AATAAGATGA	1680
	CTAGTAATAT AATTGCTATG ATTAACCACA TCATCTACTC CGACACCTCC GCCCTCATCA	1740
	AATCAGACTG ATCACTCAAC TTGCGAAGT CACTTGCGGC CTCTACATCA TCATTAGCCG	1800
35	TCATCATAAT ATATACTTGC TCAGTTACAT ACTTACCTAA CTCATACATC GCTAGTAAGA	1860
	ATAATAGTCT CAAAATTTCT TTAACCACCA CTAAACACCC CATGTTAATT TATCGATAAT	1920
40	TTGTATAGCT TGTTTAAATG CGTCTCTTTT TTCTTTGATA TCTCTATTAT CGCCATCTTC	1980
	ATCAGCTGAC ATTAACCTCAC TGTCATATTC ATATAATAGT TCTGATATTT CATTACTAGC	2040
	TACTACTAAT AAGTTTTCAT CTACATCAAT CGTTACCGTT TTCTTTGGCA TCTCCATCTC	2100
45	TCCTTATCTT AACTTGTGCC TCGTATTTGC GCTCAGCTTC TTCTTTACTC TCTGCCTCAA	2160
	CAACTGTAAA CGTCTGATTA TCTCTAGCAG TAGTAAATG TTCATGTGGT TGTCTGTGTG	2220
	AATCTTTGAA TGTTGTGACT AAGTATTGCG TCACTTCTTA TCACTCCTTT GAATGATTCT	2280
50	AAGTTTTTCT ACGAATAAAA GTATTAGTAC AACACTCAAT GTAGCCAACA TATTTTTTTG	2340
	CTTTGCAAAA TCTACTATAA CGATTAAGAC TAATAACATT CCAATTCTGC ATGTAAATAA	2400

55

	TACAAGTATT GGAAC TAATGATGTA ACTCACTTCC CCAAAACCTC CTTGACTCGA	2520
	TCTAAGATGT CTTTACACTC CGCTACTTCC GAAGCCTTTT TCTCCACGTT CTGAAACACT	2580
5	TTCGAATTCC TCCACTTGCT TTAGTTCAGG TGTCCATATA GGCACGATAA CCAATTGAGC	2640
	TAGTTTGTCT CCTTCGTTGA TTTGATAAGT TCCGTATTGT CTTATGGCGT CACTCAAATC	2700
	GATTTCTCCT TTAATATCAA AAACACCTGG TGTGATATAA CCATTCGATG CAATAGCGTC	2760
10	ATTCTTGATA TTAATCCCTA AATTGCCGTG ATATCCCGCG TCTATCTTGC CTGTTTCAAT	2820
	CACTAAATGC GTTTTACTAC TTACACCACT ACGGCTAGTT AATAGTCCGA CATAGCCCTC	2880
	TGGTATGCTT ACAGCTACAT CTGTTTTAAT CACTGCCTTT TCTTGTGGCT CAAGTACGAC	2940
15	AGTTTCAGCT GAGAATATGT CATAACCTGC ATCCGTCTTA TGATTTGCTT CGGGCAATTCT	3000
	AGCATTTTCT GATAATAGCC TTACTTGTA TGTGTTAGTC ATTTTCCTGC TCCTCCCTAG	3060
20	CTGTAGCAAA CGCTATTCTC AATTTCAATC TTTCAACAAT ATGAATTAGT GCGGTATTGA	3120
	GGAATATTTC AAATTCTTCA ATGTTCTCAT CTATAAAATC AAGTATTTCT TCCTCTTGTT	3180
	CACTGTCAAA CTCGCTTAGT ACATCCCAA TATTTATGTC GCTTTTGCTC GTTTCTAATA	3240
25	CTCTTTTGAT TATTTCTGAA TTACTTTTAT TACTCATTTT CTTGTTCCTT CCTCATATTT	3300
	ATAGACAACT TGACCTGCCA TAATCCCTAC TGCTTCATCA AGTTCAATAC CTTCTTTAAC	3360
	TGAATGTTGA ATAGCATTTG TCATTCCCTC AAGTATTTCA TCAAACGCTT GTGCTCTCTT	3420
30	ATACACGTCC TCAATCTCTT TTAGTAATCC CTCTGTGTCA TTACCGTTAT ACGCACTAGC	3480
	ACTGATCACT GATTGTTCAA TTTGTTGCGG GTTATTCATC ATTTCCATCT CCTCTAAAT	3540
	AAAGTTAGTT GCTTCTGCTC CTCGTATTCC AAACCATGTT GCTTTATATA TGTTTCGAGC	3600
35	TCTTCCGCTG TATCAAATGT CTTTTTCACG CCTTGCCAAC CTGGCAGGAT ATGCCCATGa	3660
	AAGTAATAAG TGCCGTTTAC TACATGGATA TGTGCCACTC GTTCGTTATC CTGATACAGA	3720
	TATCTCTTAG ATCCGAAAAA TTGGTTTAAG TATTCCTTAC ATGCGCTATC GGTTTTAGGC	3780
40	ATTTATGCTT CCTGCCATTT CTAAACATT TGGTTATAAG TAGTATCAAA CCAGTACGGA	3840
	TCACGTGAAT GTTTTTGAGG CACATTAAAC AAATGTGGCT TCTTCTTACG TAGTTCAGCC	3900
45	TCTTTACGTC GTTGCTAGC CATTTACGCG TCTTTGCTCT CTCGCTCCAT GATTTTGGAT	3960
	AACACAATTT CTTTATACTC AGCTAAGCGC ATACCATAAG GTGCATGTAA GGCTTCTAAC	4020
	AACGCCAGC CACCTCGTAC TCTTTTGTCA ACCATTCTTG GAGTTAAACC GTTCTTTTTT	4080
50	ATCAATTCAT TTTATGTTT GGTAAATTTA TATGGTTTAc CGTTAATCTT TACGATACTC	4140
	ATTTATTCCA CCTCTATACA TTTACTTTTT TTAATCCAAT CCTCTAATTT GTGCGTGTG	4200

55

EP 0 786 519 A2

	ACATTTAAGT TAACCATCTC AGCTTTTCCG TTTTATATC CACTAATAGT TGATCTTGAT	4320
	ACGCCAGTTT CATTGTGCAA ATCTTGGACA CTTACGTTAT CTCTAGCCAT GATTACCCCTT	4380
5	AAATTAGTTG CGAATACTcC GTTCAACTTC ATTTATTCCA CCTCTATATA TGCATGTCTT	4440
	ATTGTTATGT TGTCACTCT TAGTAATTCG TCCGGATTGT CATCTAAGCG CTTTGCCAGC	4500
	GTATCTTTTT CTTTATCCAC ATCATCGTAA TGCTGATATT CAACTTCTGT AGGTATTCTT	4560
10	ATATCAATCG TTGCGTTTAT ATATGCTTGT TGTTCATTGA GATCACTTCA TTTCTCTTTT	4620
	TCCTTTACGT CTGACTTTCA CTAAGTCCTC ATATAACCATC CATTCTTGAC CTGTGTATTT	4680
	AGGCGCTTTA CATATCCACG TTAAATTCAC ATCTCTATAC TGATATCTGA ATATCTTCGC	4740
15	TTTGATGTTG GCAACTTCAG TCGCCTTACC TTAAACGCTC ATAACCTCAA CCAGTTTCCC	4800
	TTCCTTCCAC AAAGAGAAAT CGGCTATATA CGTAATCGGT CTTTGTTTCC CGAATTTAGG	4860
20	TTGTAATTCA AATTTGGTGT GTATTTTCGAT ACGATCATAG TTAGTGCCAT TCATATTACT	4920
	TTCTAAATAT TGGAATATT CGCACTCTAC TTTGCTATCA AATACAATTC CTTTGTACTC	4980
	AACTTTCTTA GCATTGTATT TACTCAITGT GCCACCTCTA AATATCAAAT ATCGTTGCTT	5040
25	GCAATCCTAG CTCTTGCTCA TATAGAAGCC CGTGAGCGCC TTTGAATCGT TTTAGGTCAC	5100
	TATCAGTCAT AATTTTCTTT TCGTCGCTGA AATGGGCTCC TGTGAGCGAA TAAACTTCAT	5160
	TTACGTTGTC TTTATACTTG ATGACCTTAA TATCTTCCGT GCCATCTTCT CGGTATAAGT	5220
30	AATATTTTTC TTTGGCATT TTTTAACACT CCTTAATGTG TGTTTCTTC CAGTTGATTT	5280
	CATTCAATGAT TTTCTTTTCA ACTCTGTCGT AATCATCGAA AGGCGATAAC TCGTTATTGT	5340
	CCAACAATCT ATTGACCGCC CAACCAGTCT CGATATATAC ATTTGCTACA ATCGGGTCGC	5400
35	TTTGCTTTGT CTCTTCATAC ATCGATTTCA ATAAGCTTTT GAATTGCATT ATATTCATGT	5460
	GAAATACCTC TGAGTCTTCT TGTAATACTC AAATTCAATT ATTCCGGTTT CGCCGTCTTT	5520
40	GTTTITGGCT ATGTTACATT CAACAATAGA TTTGCCAGTG ATACTGTCAT CTTGTCACG	5580
	GTTATAATAA TCATCACGGT AAAGTAGCAT CGCTAACTC GCATCTGCTT CTATTCCGCC	5640
	TGATTCCTTC ATGTCCGATA GCATTGGTCT TTTATCCTGT CTAGACTCGA CACCACGATT	5700
45	CAGTTGTGAA AGTAGTACGA TGATTGCGCC TGTCTCGTTA GCGATTATCT TTAAGTCACG	5760
	TGATATCTTT TCTACTGCTA CACGTCTATC AACTTTTCGA TCAGTATCCA TCAGTTGAAG	5820
	ATAATCTATA AAAATAACTT GTTGCTGTG TGAATGCCTC ATTGTGCGC TCGCACATCT	5880
50	TGCGGTGTGA TATTACTTTT ATCAGAAATA TCGATGCCTA ATTTCAATGAT TTTATCCATC	5940
	GCAATCGTTA ACTTTGTTAA GTCATCCGGC GTTAAGTTCC TGATTTCTTT TATCTTTGTT	6000

55

	AGACTAAAGA AAGATGTTTT GTATCCATTT TGTGCTATGT TCAGCATCAT GTTTAATGCA	6120
	AAACCTGTCT TACCCACTGA GGGACGCGCT GCGATGACGA TTAATTGTGA TGGTTCTAAT	6180
5	CCCCCTATTT TGTAATCCAT TAGCTTGTA CCCGTCTTAA TTTGCTTCTT AGGGCTATCG	6240
	CTGTATAACT CTTCGACAAA CTCCTCAACA AACTTCTTGG TTCCATCTTC TTTTTTGTTA	6300
	GTAATTGTTT TTAAATCCTT GAGTTCATCA ATCAAGTTGT TAAAGTTTTG GTTCGTAGGT	6360
10	TGTTGTTTGA ACTCAGTTAC CAATTCGTTA GCTTTGTTGA GCTGATAACT TTCCAATAAT	6420
	TCTTGTTGAT AACGTTCAAA GAAGCCATAT CCAATGAAAT CGGAGTTGTA AAGTTTAGTT	6480
15	ATAGTATCTG CATCTAAAAA TTCTTTATCT TTAGTTGCTT TTAAATAGAT TTCTTGATGA	6540
	TCTATCTTTC CGACGTCCAT TACATAATTG AAAAAGGTTT TAAACTTTTC GTTCGTAAAC	6600
	ATGTAATCTT TAACTCTTAT CTTTTCTAAT ACGTCCGGTT GTTTAAGTAG CGTAGCGATT	6660
20	ATTGTACTTT CAATTTGAA TTGTCCGTAA TTCATTGCTT TTCGCCCCCA AATTCTGCCA	6720
	ACTTATTCAT GAACTTATCT AGCGCTATTT TTCTTTGTCT GACATATTCG GGGTCATTCT	6780
	GCATTTTCCA TTGGTGTGTA GCGGTTTCGT TATCTACTGG CTCGATAGAT ACTTTTTTAG	6840
25	GTTCCTTACG CATGATTGCT GGTAAGTTAG GCGGGTACGG GTTGTTACTG TTGATATAAA	6900
	CATCTACCGC TTTTACAGTT GGTTGATAAT CTCCATTTTG ACTTAATACA TCAATCCACA	6960
	TTTCTAACTT CGGTTTATCA AAATCAATGT TGTATACGTA CCTAACTTTT TTAATAATTT	7020
30	CTAATGCTTG TGTTTTGCTC ATCGGCATTA GTCATCACTC AATTCTTTTT CCATTTGTGC	7080
	AATGACATCA TCAGTAGTAT TTTTCTAGG TGCTATTTTA TTTTCTGCAT CTCTTTTGT	7140
	TTTGACATTC TCTTTAGCCC AGTTGTTTAA AACTTTAATT AAATAGCCAC CATGCGCACT	7200
35	TTTGCTTTTA GTGTACTCAA CACCTACTTT TACAACTTCA AAAGCGTTTG TACCTATATC	7260
	ATCAATAGCA AACCCTAATT GTTCCATTTG ATTAGGTGTT AACTTATCAT CCAAATTTGC	7320
40	AATTATATAT TTTATTGAAG ATGAGAAGAC GGCTTCTCTT TCTTCTTCTT TATTCTTATA	7380
	TTCTTCTTCT TTTTCTTCTT CTCTTCTTCT TTCTTCTTCT GTATCGTTAC GTAACGTTAC	7440
	GGTAACGTTA CGTTTTGCTT CTAGTAACTT TTTCTGTTTC TCACGATAGC GTTGTGTGCG	7500
45	CAATTTATTT TTTTCTTTAT GCTTAGCTTT GCTATCTAAG CTTTGATGCT TCTCCCAGTT	7560
	TGTCACTTTT ATGACACCAT TAACTTTTTT AATCATGCCC AATGTCTCAA AAGTTTGAAT	7620
	TGCTAACCTT ATTGAGTTAA TAGGTCTATT AAATTCATTT GCTAACATTT CTTCGTTGTA	7680
50	CGGCAAGTTT TCGGATAGCA TAATATAACC TTGTTTCATTG TACTTTCCTG ATAAAGTTAG	7740
	TAACTTAACC CAAATAGTTA TGATCGTATC TCTTTCGGGT AAAGCTTCGA TATATTTGAT	7800

55

	CTCCTTTTCAG	CATTTTGTG	AGCCTCTCAT	CAACTTTTAT	CCACGAGTCA	TGCAAGTGAT	7920
	ATTTATCATC	AAACGACTTA	ACGCCAATTG	CGTGCTGTTC	ATTATGATGT	TGTCTACACA	7980
5	GTGCTAACAC	ATGTTTGTG	TAGTGATTCA	TTTTGTTTCT	GTTTCATGCCT	CTGCCGACTG	8040
	CTTCATAATG	TGCCAGGTCT	GCGTGAGGCT	TTCCGCATAT	TACACAGTTG	CGGTTGATTG	8100
	TAGCCCAATA	TAATAACGCT	TTATCTTCGC	TTAACAACCT	ACTCGTTTCT	ACACTCATAG	8160
10	GTATTTGATG	ATGAAACATA	AACGCTATAA	TCAGTTCTAT	TAACTCCCTT	GCAACTTTCA	8220
	TAGAACAGTC	GCGCAGACTG	ATTTCTTCAT	AACCTTTTCAT	AATTTCCAAT	TCTGTTTGTA	8280
15	ATAATTTTCT	AGTTGATTCT	ACTGGTTTCG	CCCAGTGAAG	TTCTATATCT	CTACACATTG	8340
	CGAATATTTT	TTTGCGTTGT	TCTATAGATA	GTTTTTTTAT	GTCCGGAACC	TCTACTTCTG	8400
	CTTTTAGTGG	ATATCCGTTT	TCTAGTAAGT	CAATGTGACT	TTGTTCAAGT	TCAACACCAG	8460
20	TAGCAACGAC	GGAATAAGTA	CCGTCATTGT	CTTTCTGGTA	TCTTGTAATG	TATTGCATTT	8520
	AAACCACGTC	CTAGAACGGT	AAATCATCAT	CATTGATTTT	TATTGGACCA	TTAGCATTAG	8580
	CGAATGGGTT	TGATTGTTGA	CTCATTGGCG	TCTGTTTCCC	ATTTGCTTGC	TGTTCTTTTT	8640
25	GTTTCATCTC	ATCAGTTTTA	GGTCTGGTT	TATTAACACT	TTCATCGTCT	TTATTCCAAA	8700
	CTTTTACATA	TGAGAGTCTT	ACAAAATACT	TGCCTTGTTT	CTCGTTAAAT	TTATTTTAA	8760
	GTACAATAGT	TCCGATTTTG	TTAATTAATT	GATCTGTGTC	AAAAGTTAAA	TCTGGTAAGT	8820
30	TCAATTTAAT	TCCTAATCTA	CTAAGTAACT	CGATATATTG	TTTTTCTTGA	TAATCTTGTT	8880
	GGAATGGTGG	GACGAATTGG	TTGTGTTTGT	ATTGTTTACC	TTCGTTGTTT	TCAAAAACAA	8940
	TCGTGAAGTA	TCTGTTTCTT	CTGTCGTTAA	ACTCGACATT	TGCAACTTTT	ACTGTAAATT	9000
35	CTCCAGCTCC	TAAAAAGTCC	CCACCTTTCA	TGAATGCCTC	TTGATTAGTT	TCTTGAATGT	9060
	ATTGTTGTTCT	ACCAGTGATT	TTCATAATTT	TTATACCGTC	CTTTTAATTA	ATTTTAAATT	9120
40	ACCATTTCTA	ATTGCTTGTA	CAACATCGTT	AATACTTGA	TTAATGAAAC	GTTTGTGTT	9180
	AATTTTGATG	TTGCTTGAGT	GTCTTATCTT	TGTCTCGAAT	AAATTTGATG	GTTCAGCGTT	9240
	AAGTACATAT	TGATAAGTTT	TTTCGCCGTC	TTGCTCATGT	TCTTCTATTG	TCATTCTTGC	9300
45	TAACACGTCA	GATTGACTGA	TGACTGCTTT	TTTTATTTGG	TCTTGTGCCT	CTATCGTGAT	9360
	TGTTGGATTG	ATAGTACTTC	CCTCATCATC	TTGTCTTTG	TTAATGCCCT	CGTGTCCGCT	9420
	TATAGCAAGA	TGAAATTGAT	AATGTTCTTG	TAATTTAGAA	ATATAACGAT	AAATACTTAC	9480
50	AATGCGTGTA	GCACACTCGC	CCCAATCATT	AAATGTCGGT	TTCTTTGATT	TACCGTCCAT	9540
	GATGTCGTCC	ATAGTGATAT	CACGTAACCT	TTGGATTGTT	TCAATCACTA	CAACATCAAT	9600

55

AAAATGCTTA TAATTCTTAA TCTGCACAAC TGCCCCATCT TCTGTTACCG TTGTTCCGTC 9720  
 CTCATTTATA TCTAGTACTA AGGCATTGTT ATCTTTTGTT AAAAACGTAG TTTTACCAGT 9780  
 5 ACCGAACTTG CCGTATATCG CAAATTTATA AACTTGTTT GCATTTTGTT TGCTGATGTC 9840  
 TTTTACACCT AGTTGCGTTA AAATATCGAC ATCTTGATTA GTTTTTTCAG TCATCTATTC 9900  
 TCCACCTTT ACCGTGTATG ACGTTGGTTT CTCCACAATG CTAGCACCTT CTAAAACCTC 9960  
 10 GCCGTTTTCG TCAATCAATG TGCCGTTTTC AGTTACATTG AAATCTTTCT TAATGTCTGA 10020  
 TTGGCTAAGT TTTTTAGTTA CTTTACATA GTTGTCAAAA CCTCGTTGCT CAAGTTGTTT 10080  
 AATGACTTCT TGCTCATTGC TAACTTGAAT GACTTTTGAA CCTTTTCTGG CTGTCACCTT 10140  
 15 TCCGTAAGTG TATTCAACTT GAATTGCTA TCTTGTCTT TTTGTATTCT GTAATATTCA 10200  
 ATTACAAGGC TTTGTAAATA TTCTTTGCCA CTCTGTAATT TTTCTACTTC TTTATCTTTC 10260  
 CATTGTTTTA TGC GTTCAAT TTCTTTATTT GCTAAATCGT TGATTTTATT CTCTTTAGTT 10320  
 GTGATTGCAT CCAGTTTCTT AAAAACCCAG TTAGCACTGT CTAGATCAGT nACTTTGAAT 10380  
 CGGTCGTCTT GTTCGAATGT n 10401

25 (2) INFORMATION FOR SEQ ID NO: 150:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2989 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

35 TTTCTCTCTA TTATTCTCGA TGC GTAGATA ATTGTTTAAA TTTAAGTTTA TAGTAATGTT 60  
 GAGTTTATAA TTTTATATAT CTAAAAACAG GTGTTGTATA TATAATCATT CATCTAGTTA 120  
 TACTTACTTT AAAAATAATA TAATTTCATG CGATGCAATT CATTGATGGA TGTTTTTAAT 180  
 40 CTTAATCAAA TCCAAATAAA GCATATATTT TTAAATTCAC TTTCTTTTGA ATCGATTTTT 240  
 ATCTCTTGnA TTAACTTTT CCATTGTTTC ATTAAAGCTC TCTGTCATAT CTATTCCCAT 300  
 TGAATTCGCT AACATAACA ACACAAATAA ATTATCACCT AATTCTGCTT TAATCGTATT 360  
 45 TGCTTCCTCT GAATCTTTCT TCTTTTTTTC ACCATAGGTA TGATTTATTT CACGTGCAAG 420  
 TTCGCCCCACT TCTTCAGTCA ATCTAGCTAA GTTAGCTAAT GGTGAAAAAT ATCCTGTTTT 480  
 50 AAATGTCCA ATATATTCAT CAACTTCACG TTGCATTTCT ACCATTGATT TCATTTCTAC 540  
 GTTCTCCTTA TATTGCATTT CTAATATAGT ATATATCAAT TTGAAGTCTC ATGCATGTTT 600

55

	AATTCAGTTT ATATAAATGT AATGCATTCC TAACTAAATT AAATCAATTG AAATTGGGAT	720
	TATAACTTTA TGATACGTAC CACTACAATA AAATAATATA GTGAATAATC TACCATTAGA	780
5	AAAATAAGCA CAAAAAACT AGCAACCACA CAAAAATGTG ATTAGCTAGT TAATAAGTGT	840
	CTAATTTAAG TTAATTGTTA ATCTATAAGA TTAATCACTT GAACGCGCAA TCAAAATAAT	900
	ACGTACAAGC TCTGCTACAG CGACTGCAGT TGCTGCAACA TAAGTCATTG CTGCTGCAGA	960
10	TAATACTTTA CGCGCATGCT TGTATTCTTT TTCATTTACA ATGTTCAATG CCGTAATTG	1020
	TTTCATCGCT CTTGAACTCG CATCAAATC AACTGGTAAC GTAACAATTG AGAATAATAC	1080
15	CGCTAATGAC ATTAAACCAG CACCAATCCA TAAAGCAGTT GAACCAaATG CACTACCTAT	1140
	CGCTGTTAAG ATAATACCTA ACATGATGAT CATATAACTT AATGAACTCC CTAGGTTTGC	1200
	AACAGGTAAT AATGCTGCTC TGAATCTTAA GAACCAATAT CCTTGGTGAT CTTGAATGGC	1260
20	ATGACCAACT TCGTGGGCTG CAATTGCAGT TCCAGCAACT GATGGTCTGT CATAGTTTGC	1320
	AGGAGATAGT GAAACAACCT TCTTTTTAGG ATCGTAATGA TCTGTTAAGA ATCCTTCACC	1380
	TTTAACAACCT TCGACATCAT AAATACCGTT TGCATGTAAA ATTTCTAATG CAACTTCACG	1440
25	ACCCGTTTTA CCACTAGTTG ATCTAACTG TGAATATTTT TCATAGTTAG ATTTAACTTT	1500
	GTGTTGTGCC CATAAAGGAA GCACCATTAA TATTACGAAA TAAATTATCA TAGTAAAAAT	1560
	TGAAGACAAT AACTCACTC TCCTTTATAA ATATTTTACT GTCATTTGCC GTTTTTATCA	1620
30	AATCATTTAC ACTTTAATAA TTGTTTTAAT TCAATATAAA GCAAAAGTCC AAAAACACTT	1680
	AGACAACATG ATAATACACC AATTTGCCAC ACATGTGTAG TTATAAAATC ATAATATGGA	1740
	AATTGAAGGT GAAAATAGTC AATATAATCA TTCAAAAACA CCCAAATCAT yGCTACACTG	1800
35	ATTCCAATCA TAGAACGTTT AAACCTAGGA TAGAAGTAAA TTGCCTGAAC AGCCATTATA	1860
	CTGTGGGAAA ACATTAATAC CAAACCATT ACTGTAATAT CACCTTGTTT AATAATAAAT	1920
40	AATATATTCA TTATAACTGC CCAAATCCCA TATTTGAATA ATGTTACAAA TGCCAGTGCA	1980
	TCGATAATAC TATTTTGTGTT TTGAATTAAT ATCAATGAGA TAGAAATAAC TAAGTATAAT	2040
	ATTGCAGTTG GGCTATCTGG AACAAAAATC TTAAATGCC AGGGCGTATG ACTTAATTGT	2100
45	TCACCATACC ATATATAACC ATAAATCATC CCTAATATAT TACAAATGAG TAGCATCATT	2160
	AACCAAGAAC GTTGATAAAG TGTATATTGC CAAAATGCTT TAATTGTCAT CTGCTAAGTC	2220
	CTCAAATTGA TTATGTTTAT TTAGTAGCTT GAGTGTATTT AAAATTGCG TTAGTTGATA	2280
50	AAAACGTTGC TTTTCATTCA TCTGTAACT TAAATCAATA TTGTGTAACA AGTAATCTAT	2340
	TAATAACGCA TGTTTATGCC GATCTATAGC CATACTATTT AAGTCATGAA GATAAGTTTG	2400

55

TGACACGTTT GCGAAGTGAA TTTGAATATC AAAAGCACAG TTATGATTAG CGATATAATC 2520  
 AAATATTTCA TTGTATTCA TTAACCTTAT ATTACGCTTA GTAAATTGAA TTGCAGAAGC 2580  
 5 GTGACTTCCC ACTTCTGCAA TTTCTAATGT TTCATGATGA TTAATTTTTG TATCTACAAA 2640  
 ATGAATGTTT GCCAATTTTCG CCTCATTAC TTTTATATAG TTAAGCACCC AAACGTCAAT 2700  
 ACGCGACTTA AATCGATATT GAAAAAGTAA ATATTCAATA AAACCTTCTT TAATTTGATT 2760  
 10 GAGTGTCTCT GACATCAAAT ACCCCATTTT AAGATTGCAA TCTTGaTAAT TCGTCATGCC 2820  
 AATTTTCGTT ACTTGGcTCT AGTTCCAACA ATTGATTTAA AATAGTAATT GCTTGTTCCT 2880  
 TTTGACCAAT TTCAATTAAA TAGAAATAAT AATCACTCAT AAAATCAATA TTTGTTTTCA 2940  
 15 TCGTTGGATA TGCTAATTCA AAGAAATGTT GAGCTTCTTT ATCTCGCTC 2989

(2) INFORMATION FOR SEQ ID NO: 151:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1143 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

CATCAACTCC TTAATTACAC TGTAATGAT ATGCGTCTTT TTGACAACTA TATTGTCAA 60  
 30 ATCTACACCA AAAAATATGA TTATCCACCT ATGTATGACA TTTTGAAACA AACACCTCAA 120  
 CGCCTACAAG TCATAATTGT TACTTTTCGT TACACCTTCC TGCATAATTA ACAGCATTCT 180  
 AATTTTAGTA TGATGCACGC ATTTTCACTA AATCAAACCA TTCAAAGGAG ACTATTATGG 240  
 35 CATTTACATT ATCTGCAATT CAACAAGCAC ATCAACAATT TACTGGTGTT GACTTTCCAA 300  
 AACTATTCAA AGCTTTTAAA GATATGGGGA TGAATTACAA TATCGTCAAC ATTCAAGATG 360  
 GCACTGCAAC ATACGTACAT CAATCAGAAG ATGATATCGT TACGTCTCTT GTAAAAAGTA 420  
 40 ATCATCCTGT TGCTCAAAAA TCAAACAAAA CAATAGTTCA AGACGTCTTA ACTAGACATC 480  
 AACAAAGGGA AACAGATTTT GAAACATTTT GTGATGAAAT GGCTGAAGCT GGCATTTATA 540  
 45 AATGGCATAT CGATATTcma GCGGGCACTT GTACTTATAT CGACTTGCAA GACCAAGCTG 600  
 TTATTTTACA ATTAATCCCT CAATAAACTA TATTTATAGC AACATTTTAA TTATTTTATA 660  
 AAATTTTATT GATAATCATT ATCGTTCGGT ATAAAGTAAA TACTATATAC TACTTATGAG 720  
 50 TGAGGTTGAT TATCATGATA ACTAACCTT TTATTTTAGG CATCACAGGC CCAACAAGTC 780  
 TTGTCGTCAT TAGCATTATC GCTTTAATTA TTTTGTGTC GAAAAAATTA CCACAATTTG 840

55



AGTCTCACGA TACACCCAGT AAGGAATCGA AACAACAGCG AGAGCAATAG CACTGACCAC 960  
 ACCTTACTGG TTCACCTT TAG CGAACTACGC CATCGGTTAG TAAAAATTTT ATTGTCGTTT 1020  
 5 GTCATTACGG TCATCGTCGT ATATGTYTCA TCATTTTGGT GGATGACACC ATTCATAACG 1080  
 TATATyACCC GgCACATGTG TcCTTACATG CATTTcATTc ACAGAAATGA TACAAATAAC 1140  
 GTG 1143

10

(2) INFORMATION FOR SEQ ID NO: 152:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7953 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

CAACGCCTGA ACGTAAACCA TATCGTTTTCG CGATTTCTCTC ATCTTGACTA TTTACTAAAA 60  
 ACTCTCTCAT GGCGATTAAT GTTCTTTTTT CTTCTTTAGT TAATGGTAAT TCTAACTCAG 120  
 25 CTGCTTTTTG ACGCAAAGTT GGATGACCAT CTCTAATGAT GTCTTTTCATT GTTAACATAT 180  
 ATTGCACCTT CCTTATTTTA ATTTGTTTTA GTTGAATGAC AGTAAAAAGG TTGTTAAGAT 240  
 ACTCATACAT TTTTATGTGT AAATATCTAC AAAGTTAACC AACTACTGCC AATGTTTATT 300  
 30 TTAGATAGTA TATGTAAATT TTCAaGaTAT GCgTAATTGC gTTAAAAAAT GaTTAAAGTG 360  
 TTGGTTTCAA GCAATGaTAC TTTAGAAATT TATTTATCAT CTTGACTTTA AAAATTATAT 420  
 TATAAATGAC GTAACGTGCA ACAGATATAC TTAGTArTGA AGATGTGTAA TGTAATTGTT 480  
 35 TAA AATTGAT TTCCAAGCAG ATTTTATTTA TCATTTAATT TAAATAGCAA GTGGAGGTAC 540  
 AAGTAATGAA ATTTGGAAAA ACAATCGCAG TAGTATTAGC ATCTAGTGTC TTGCTTGCAG 600  
 40 GATGTACTAC GGATAAAAAA GAAATTAAGG CATATTTAAA GCAAGTGGAT AAAATTAAAG 660  
 ATGATGAAGA ACCAATTAAA ACTGTTGGTA AGAAAATTGC TGAATTAGAT GAGAAAAAGA 720  
 AAAAATTAAC TGAAGATGTC AATAGTAAAG ATACAGCAGT TCGCGGTAAA GCAGTAAAGG 780  
 45 ATTTAATTAA AAATGCCGAT GATCGTCTAA AGGAATTTGA AAAAGAAGAA GACGCAATTA 840  
 AGAAGTCTGA ACAAGACTTT AAGAAAGCAA AAAGTCACGT TGATAACATT GATAATGATG 900  
 TTAAACGTAA AGAAGTAAAA CAATTAGATG ATGTATTAAA AGAAAAATAT AAGTTACACA 960  
 50 GTGATTACGC GAAAGCAtaT AAAAAGGCTG TAAACTCAGA GAAAACATTA TTTAAATATT 1020  
 TAAATCAAAA TGACGCGACA CAACAAGGTG TTAACGAAAA ATCAwAAGCA ATAGAACAGA 1080

55

	AAGAAAAGCA AGACGTTGAT CAATTTAAAT AATTAATATA ATACAGATGG TAGGAAACAA	1200
	CTAATACAGT TCCTATTATC TGTATCTTTT TTTATTAAAA CAGAACTTTT TCAAATGGTT	1260
5	TAACAGTCCC ATTTATTGT GGTACAATTA GTAAGGATAA AATGAATTC TATACAATTA	1320
	TGGGAAAGGT ATTGTGAATT GAATGGCTCC TAAGTTACAA GCCCAATTCG ATGCAGTAAA	1380
	AGTTTTAAAT GATACTCAAT CGAAATTTGA AATGGTTCAA ATTTTGGATG AGAATGGTAA	1440
10	CGTCGTAAAT GAAGACTTAG TACCTGATCT TACGGATGAA CAATTAGTGG AATTAATGGA	1500
	AAGAATGGTA TGGACTCGTA TCCTTGATCA ACGTTCTATC TCATTAAACA GACAAGGACG	1560
15	TTTAGGTTTC TATGCACCAA CTGCTGGTCA AGAAGCATCA CAATTAGCGT CACAATACGC	1620
	TTTAGAAAAA GAAGATTACA TTTTACCGGG ATACAGAGAT GTTCTCAA TTTTGGCA	1680
	TGGTTTACCA TTAAGTGAAG CTTTCTTATT CTCAAGAGGT CACTTCAAAG GAAATCAATT	1740
20	CCCTGAAGGC GTTAATGCAT TAAGCCCACA AATTATTATC GGTGCACAAT ACATTCAAGC	1800
	TGCTGGTGTT GCATTTGCAC TTAAGAAACG TGGTAAAAAT GCAGTTGCAA TCACTTACAC	1860
	TGGTGACGGT GGTCTTTCAC AAGGTGATTT CTACGAAGGT ATTAACCTTG CAGCAGCTTA	1920
25	TAAAGCACCT GCAATTTTCG TTATTCAAAA CAATAACTAT GCAATTTCAA CACCAAGAAG	1980
	CAAGCAAACCT GCTGCTGAAA CATTAGCTCA AAAAGCAATT GCTGTAGGTA TTCCTGGTAT	2040
	CCAAGTTGAT GGTATGGATG CGTTAgcTGT nATATCAAGC AACTAAAGAA GCACGTGACC	2100
30	GCGCAgTTGC AGGTGAAGGT CCAACATTAA TTGAACTAT GACATATCGT TATGGTCCTC	2160
	ATACAATGGC TGGTGACGAT CCAACTCGTT ACAGAACTTC AGACGAAGAT GCTGAATGGG	2220
	AGAAAAAGA CCCATTAGTA CGTTTCCGTA AATTCCTTGA AAACAAAGGT TTATGGAATG	2280
35	AAGACAAAGA AAATGAAGTT ATTGAACGTG CAAAAGCTGA TATTAAAGCA GCAATTAAAG	2340
	AGGCTGATAA CACTGAAAAA CAACTGTTA CTTCTCTAAT GGAAATTATG TATGAAGATA	2400
40	TGCCTCAAAA CTTAGCAGAA CAATATGAAA TTTACAAAGA GAAGGAGTCG AAGTAAGCCA	2460
	TGGCACAAAT GACAATGGTT CAAGCGATTA ATGATGCGCT TAAACTGAA CTTAAAAATG	2520
	ACCAAGATGT TTAAATTTTT GGTGAAGACG TTGGTGTTAA CGGCGGTGTT TTCCGTGTTA	2580
45	CTGAAGGACT ACAAAAAGAA TTTGGTGAAG ATAGAGTATT CGATACACCT TTAGCTGAAT	2640
	CAGGTATTGG TGGTTTAGCG ATGGGTCTTG CAGTTGAAGG ATTCCGTCCG GTTATGGAAG	2700
	TACAATTCTT AGGTTTCGTA TTCGAAGTAT TTGATGCGAT TGCTGGACAA ATTGCACGTA	2760
50	CTCGTTTCCG TTCAGGCGGT ACTAAACTG CACCTGTAAC AATTCGTAGC CCATTTGGTG	2820
	GTGCGGTACA CACACCAGAA TTACACGCAG ATAACCTAGA AGGTATTTTA GCTCAATCTC	2880

55

	CTATTAGAAG TAATGACCCA GTCGTATACT TAGAGCATAT GAAATTGTAT CGTTCATTCC	3000
	GTGAAGAAGT ACCTGAAGAA GAATATACAA TTGACATTGG TAAGGCTAAT GTGAAAAAAG	3060
5	AAGGTAATGA CATTTCAATC ATCACATACG GTGCAATGGT TCAAGAATCA ATGAAAGCTG	3120
	CAGAAGAACT TGAAAAAGAT GGTATTCTG TTGAAGTAAT TGACTTACGT ACTGTTCAAC	3180
	CAATCGATGT TGACACAATT GTAGCTTCAG TTGAAAAAAC TGGTCGTGCA GTTGTAGTTC	3240
10	AAGAAGCACA ACGTCAAGCT GGTGTTGGTG CAGCAGTTGT AGCTGAATTA AGTGAACGTG	3300
	CAATCCTTTC ATTAGAAGCA CCTATTGGAA GAGTTGCAGC AGCAGATACA ATTTATCCAT	3360
	TCACTCAAGC TGAAATGTT TGGTTACCAA ACAAATGA CATCATCGAA AAAGCAAAAG	3420
15	AAACTTTAGA ATTTTAATAC ATTTTAAAAG TTAACGAAGT TAGCGTATTT TAGTCTCATT	3480
	GATTAAATG AAATGTTTAA TTTACGAAAT CTTAGGAGGG CAAAAACGTG GCATTTGAAT	3540
20	TTAGATTACC CGATATCGGG GAAGGTATCC ACGAAGGTGA AATTGTAAAA TGGTTTGTTA	3600
	AAGCTGGAGA TACTATTGAA GAAGACGATG TTTTAGCTGA GGTACAAAAC GATAAATCAG	3660
	TAGTAGAAAT CCCATCACCA GCATCTGGTA CTGTAGAAGA AGTTATGGTA GAAGAAGGTA	3720
25	CAGTAGCTGT AGTTGGTGAC GTTATTGTTA AAATCGATGC ACCTGATGCA GAAGATATGC	3780
	AATTTAAAGG TCATGATGAT GATTCATCAT CTAAAGAAGA ACCTGCGAAA GAGGAAGCGC	3840
	CAGcAGaGCA AGCACCTGTA GCTACTCAAA CTGAAGAAGT AGATGAAAAC AGAACTGTTA	3900
30	AAGCAATGCC TTCAGTACGT AAATACGCAC GTGAAAAAGG TGTTAACATT AAAGCAGTTT	3960
	CTGGATCTGG TAAAAATGGT CGTATTACAA AAGAAGATGT AGATGCATAC TTAAATGGTG	4020
	GTGCACCAAC AGCTTCAAAT GAATCAGCTG CTTCAGCTAC AAGTGAAGAA GTTGCTGAAA	4080
35	CTCCTGCAGC ACCTGCAGCA GTAACATTAG AAGGCGACTT CCCAGAAACA ACTGAAAAAA	4140
	TCCCTGCTAT GCGTAGAGCA ATTGCGAAAG CAATGGTTAA CTCTAAGCAT ACTGCACCTC	4200
40	ATGTAACATT AATGGATGAA ATTGATGTTT AAGCATTATG GGATCACCGT AAGAAATTTA	4260
	AAGAAATCGC AGCTGAACAA GGTACTAAGT TAACATTCTT ACCTTATGTT GTTAAAGCAC	4320
	TTGTTTCTGC ATTGAAAAAA TACCCAGCAC TTAACACTTC ATTCAATGAA GAAGCTGGTG	4380
45	AAATCGTTCA TAAACATTAC TGGAATATCG GTATTGCAGC AGACACTGAT AGAGGATTAT	4440
	TAGTACCTGT TGTTAAACAT GCTGATCGTA AGTCTATTTT CCAAATTTCA GATGAAATTA	4500
	ATGAATTAGC TGTTAAAGCA CGTGATGGTA AATTAACAGC CGATGAAATG AAAGGTGCTA	4560
50	CATGCACAAT CAGTAATATC GGTTCAGCTG GTGGACAATG GTTCACTCCA GTTATCAATC	4620
	ACCCAGAAGT AGCAATCTTA GGAATTGGCC GTATTGCTCA AAAACCTATC GTTAAAGATG	4680
55		

	ATGGTGCAAC TGGCCAAAAT GCAATGAATC ACATTAAACG TTTATTAAAT AATCCAGAAT	4800
	TATTATTAAT GGAGGGGTAA AACATGGTAG TTGGAGATTT CCCAATTGAA ACAGATACTA	4860
5	TAGTAATCGG AGCAGGTCCT GGTGGATACG TTGCAGCAAT TCGTGCAGCT CAATTAGGAC	4920
	AAAAAGTAAC AATCGTTGAG AAAGGTAATC TTGGTGGTGT TTGCTTAAAC GTAGGATGTA	4980
10	TTCCTTCAAA AGCATTACTA CATGCTTCTC ACCGTTTTGT TGAAGCACAA CATTCTGAAA	5040
	ACTTAGGTGT TATTGCTGAA AGTGTCTCTT TAACTTCCA AAAAGTTCAA GAATTCAAAT	5100
	CATCAGTTGT TAATAAATTA ACTGGTGGTG TTGAAAGCTT ACTTAAAGGT AACAAAGTTA	5160
15	ACATCGTTAA AGGTGAAGCA TATTTCTAG ATAACAATAG CTTACGTGTT ATGGACGAAA	5220
	AGAGCGCACA AACATACAAC TTTAAAAATG CAATCATTGC AACAGGTTCA AGACCAATTG	5280
	AAATTCCTAA TTTCAAATTC GGTAAACGTG TTATCGACTC AACAGGTGCT TTAAACTTAC	5340
20	AAGAAGTACC aGGTAAATTA GTTGTAGTTG GTGGAGGATA CATTGGATCA GAATTAGGTA	5400
	CAGCATTTGC TAACTTTGGT TCAGAAGTAA CCATCCTTGA AGGTGCTAAA GATATCTTAG	5460
	GTGGCTTCGA AAAACAAATG ACACAACCTG TTAAAAAAGG TATGAAAGAA AAAGGTGTTG	5520
25	AAATCGTTAC TGAAGCTATG GCTAAATCAG CTGAAGAAAC AGATAACGGA GTTAAAGTTA	5580
	CTTATGAAGC TAAAGGCGAA GAGAAAACAA TCGAAGCTGA TTATGTATTA GTAAGTGTAG	5640
	GTCGTCGTCC AAACACAGAC GAATTAGGCC TAGAAGAATT AGGTGTTAAA TTCGCTGACC	5700
30	GTGGATTATT AGAAGTTGAT AAACAAAGCC GTACGCTCTAT CAGCAATATC TATGCAATTG	5760
	GTGATATCGT TCCAGGTTTA CCACTTGCTC ACAAAGCTAG CTATGAAGCT AAAGTTGCTG	5820
35	CTGAAGCAAT TGATGGTCAA GCTGCTGAAG TTGATTACAT TGGTATGCCA GCAGTATGCT	5880
	TTACTGAACC AGAATTAGCT ACAGTTGGTT ATTCAAGAGC GCAAGCTAAA GAAGAAGGTT	5940
	TAGCAATTAA AGCTTCTAAA TTCCCATATG CAGCAAATGG TCGTGCATTA TCATTAGATG	6000
40	ATACTAACGG ATTTGTAAA CTTATTACAC TTAAAGAAGA TGATACTTTA ATCGGTGCTC	6060
	AAGTAGTTGG TACTGGTGCA TCAGATATTA TCTCTGAATT AGGTTTAGCA ATTGAAGCTG	6120
	GTATGAATGC TGAAGATATC GCATTAACAA TCCATGCACA TCCAACATTA GGTGAGATGA	6180
45	CTATGGAAGC AGCAGAAAAA GCTATCGGAT ACCCAATCCA TACAATGTAA TAACTGATTA	6240
	TCTATAAAGA TTCAGTCATT AAAAGCTGTA GCATATGCTA CGGCTTTTTT GTTTTAGGTA	6300
	AAGTAATGTA AGGAAATTGA TTTGAGATAT CGTTAACATG TGACATGCAT GTTATACTAG	6360
50	CGATGCTAAT AAAAGAATTG AAATGGAGGG TTCAACAATG GAATATGAGT ATCCAATTGA	6420
	TTTAGACTGG AGTAATGAAG AGATGATTTT AGTGATAAAT TTCTTTAATC ATGTAGAGAA	6480
55		

AATTGTGCCT GCTAAAGCAG AGGAAAAACA AATTTTAAAT ACTTTCGAAA AAAGTAGTGG 6600  
 CTATAATAGT TACAAAGCAG TTCAAGATGT AAAAAGTCAC TCTGAAGAAC AAAGAGTAAC 6660  
 5 AGCTAAAnAA TAATTCGTTC GAAATTAACA CAATTTAATA GGAATTTTTC TTTAAAACTA 6720  
 TTGCTAATAA AGCTATATTT TGATACCTTT ATCAAGTGTT AAACAAAATG TTTGATAAAA 6780  
 GTAAACTTAA TATAGCTTTT TTAGGTGGAA AAATAAATGA ACATAGGTAA TAAAATTAAA 6840  
 10 AATCTTAGAA GAATTAAAAA TTTAACGCAA GAAGAACTTG CTGAACGTAC AGACTTATCG 6900  
 AAAGGCTACA TTTCACAAAT AGAAAGTGAA CATGCCTCAC CAAGTATGGA AACTTTCTTA 6960  
 AATATTATAG AGGTGTTAGG AACGACGCCA AGTGAATTTT TTAAAGACAG TGAAAATGAA 7020  
 15 AAAGTATTAT ACAAGAAGGA AGAACAAGTT ATTTATGATG AGTATGATGA AGGTTATATA 7080  
 TTAAATTGGT TAGTTTCAA GTCAAATGAA TATGATATGG AGCCATTAAT ATTAACTTTA 7140  
 AAGCCTGGAG CATCATATAA AAATTTTAAAT CCATCAGAGT CTGATACGTT TATTTATTGT 7200  
 20 ATGTCAGGTC AGATAACACT TAATTTAGGC AAAGAGATAT ATCAAGCACA AGAAGAAGAC 7260  
 GTTTTGTATT TTAAAGCACG AGATAATCAT CGTTTGTCAA ACGAATCAAA CAATGAAACA 7320  
 25 CGAATACTTA TTGTAGCGAC AGCTTCATAT TTATAGGGGG GATCTTATTT GGAACCGTTA 7380  
 TTATCATTAA AATCAGTTAG TAAAAGCTAT GATGATCTTA ATATCTTAGA TGACATAGAT 7440  
 ATTGATATTG AATCAGGATA CTTTATACA TTATTAGGTC CTTCAGGTTG TGGTAAAACA 7500  
 30 ACAATTTTAA AATTAATTGC AGGGTTTGAA TATCCTGACA GTGGTGAAGT GATTTATCAA 7560  
 AACAAACCAA TTGGTAATTT ACCACCAAAT AAACGTAAAG TGAATACAGT CTTTCAAGAT 7620  
 TATGCATTAT TTCCACACTT AAACGTCTAT GATAATATCG CTTTGGTTT GAAATTAAAA 7680  
 35 AAATTATCAA AAACCGAAAT TGATCAAAAA GTAAGTGGG CATTAAAATT AGTAAACTT 7740  
 TCAGTTATG AAAAAAGAAA TATTAATGAA ATGAGTGGCG GACAAAAGCA ACGTGTGCA 7800  
 ATTGACAGTG CTATCGTAAA TGAACCAGAA ATATTATTGT TAGATGAATC TTTATCCGCA 7860  
 40 TTAGATTTGA AATTGCGTAC TGAAATGCAA TATGAATTAC GAGAATTGCa ATCTAGATTA 7920  
 GGTATTACAT TTATATTTGT aACACATGAT CCA 7953

45 (2) INFORMATION FOR SEQ ID NO: 153:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2347 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

55

	GGCGTGATCA TACGACCGTC ATTCATGCTC ATGAAAAAAT ATCTAAAGAT TTAAAAGAAG	60
	ATCCTATTTT TAAACAAGAA GTAGAGAATC TTGAAAAAGA AATAAGAAAT GTATAAGTAG	120
5	GAAACTTTGG GAAATGTAAT CTGTTATATA ACAGCACTAA TGATnACAAT CATTTTTTTAC	180
	ATTCTATAT GCTAATGTGG CAAGATGAGC AAAACTCATT TTGTGGATaA TGTTTaAAAG	240
	TCATACACAC CATAACAAG TTATCAACAT GTGTATAAyT cGcCAAATCT ATGTTTTTTAA	300
10	GACTTATCCA CCAATCCACA GCACCTACTA CTATTACTAA GAACTTAAAA CCTATATAAT	360
	TATATATAAA CGACTGGAAG GAGTTTTAAT TAATGATGGA ATTcACTATT AAAAGAGATT	420
15	ATTTTATTAC ACAATTaAAT GACACATTAA AAGCTATTTT ACCAAGaACA ACATTACCTA	480
	TATTAAGTGG TATCAAAATC GATGCGAAAG AACATGAAGT TATATTaACT GGTTCAGACT	540
	CTGAAATTTT AATAGAAATC ACTATTCCTA AAAGTGTAGA TGGCGAAGAT ATTGTCAATA	600
20	TTTCAGAAAC AGGCTCAGTA GTACTTCCTG GACGATTCTT TGTGTATATT ATAAAAAAAT	660
	TACCTGGTAA AGATGTTAAA TTATCTACAA ATGAACAATT CCAGACATTA ATTACATCAG	720
	GTCATTCTGA ATTTAATTTA AGTGGCTTAG ATCCAGATCA ATATCCTTTA TTACCTCAAG	780
25	TTTCTAGAGA TGACGCAATT CAATTGTCGG TAAAAGTGTCT TAAAAACGTG ATTGCACAAA	840
	CAAATTTTGC AGTGTCCaC TCAGAAACAC GCCCAGTACT AACTGGTGTG AACTGGCTTA	900
	TACAAGAAAA TGAATTAATA TGCACAGCGA CTGACTCACA CCGCTTGGCT GTAAGAAAGT	960
30	TGCAGTTAGA AGATGTTTCT GAAAACAAAA ATGTCATCAT TCCAGGTAAG GCTTTAGCTG	1020
	AATTAAATAA AATTATGTCT GACAATGAAG AAGACATTGA TATCTTCTTT GCTTCAAACC	1080
	AAGTTTTATT TAAAGTTGGA AATGTGAAGT TTATTCTCG ATTATTAGAA GGACATTATC	1140
35	CTGATACAAC ACGTTTATTC CCTGAAAAGT ATGAAATTAA ATTAAGTATA GACAATGGGG	1200
	AGTTTATCA TGCATTGAT CGTGCCTCTT TATTAGCGCG TGAAGGTGGT AATAACGTTA	1260
40	TTAAATTAAG TACAGGTGAT GACGTTGTTG AATTGTCTTC TACATCACCA GAAATTGGTA	1320
	CTGTAAAAGA AGAAGTTGAT GCAAACGATG TTGAAGGTGG TAGCCTGAAA ATTTCAATCA	1380
	ACTCTAAATA TATGATGGAT GCTTTAAAAG CAATCGATAA TGATGAGGTT GAAGTTGAAT	1440
45	TCTTCGGTAC AATGAAACCA TTTATTCTAA AACCAAAAGG TGACGACTCG GTAACGCAAT	1500
	TAATTTTACC AATCAGAACT TACTAAAAAT AAATATAAAT AAAGGATGAC GTGATTAATT	1560
	AAAACGTCAT CCTTTATTTT TTGGCAAAAA TAATTCTAGG TCGGTATGTA AAATAAATTT	1620
50	GGCAGCATTT TAAACAGCAA ATAAAAGACG CCAATTAAAT TTATGACAAA TGTATCCAAA	1680
	ATTTAATAAG TGTGCTTATA TGCCCTTTAA ATTTAAATTT TTAATAGTCA ATAACAAGTT	1740

55

AAAAATAAGA ATTAATTATT TATATGTAAA CGGTTTCTAC CTCTATTTTA AATGAAATTT 1860  
 GTGACAAAAA AAGGTATAAT ATATTAATGA CATACAAAGA AATGGAGTGA TTATTTTGTT 1920  
 5 TCAAGAAGTT GTAGTAGAAG GAGACATTAA TTTAGGTCAA TTTCTAAAAA CAGAAGGGAT 1980  
 TATTGAATCT GGTGGTCAAG CAAAATGGTT CTTGCAAGAC GTTGAAGTAT TAATTAATGG 2040  
 AGTGCCTGAA ACACGTCGCG GTAAAAAGTT AGAACATCAA GATCGTATAG ATATCCCAGA 2100  
 10 ATTACCTGAA GATGCTGGTT CTTTCTTAAT CATTCAATCAA GGTGAACAAT GAAGTTAAAT 2160  
 AACTCCAAT TAGAAAATTA TCGTAACTAT GATGAGGTTA CGTTGAAATG TCATCCTGAC 2220  
 GTGAATATCC TCATTGGAGA AAATGCACAA GGGAAAGACA AATTTACTTG GAATCAATTT 2280  
 15 ATACCTTAGC TTTAGCAAAA AGTCATAGAA CGAGTAATGG ATAAGGGACT CCATACCGTT 2340  
 TTAATGC 2347

20 (2) INFORMATION FOR SEQ ID NO: 154:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13542 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

30 ACAAGACGTh TCTATAACTT ATCTGAAATC GCTCGTCAAG ATAAAGATTA TGCAACTATC 60  
 TCATTCTTAA ACTGGTTCTT AGATGAACAA GTCGAAGAAG AATCAATGTT TGAAACTCAC 120  
 ATCAATTATT TAACTCGTAT CGGCGATGAC AGCAATGCAT TATATCTTTA CGAAAAAGAA 180  
 35 CTTGGCGCTC GTACATTCTA CGAAGAATAA TTAAACATCA CTACAATAGA CAGATAAATA 240  
 TCATACGACA TGATAGGCAT TTGGGTCACT TACAATAACC CAATGTCTAT ATTATTTTGC 300  
 TTTACGGAGA TCACTAGATT CATTTTCTGA ATCATTGATC TCGGTTTMTT CATTTTCAAG 360  
 40 GCTAATTATT GTATTTTTAG TCATTTATTT TTTAACTAC TAATGTTAAT AACTCTAAAT 420  
 TTGATGTTGA ATTAATTTGA CGATTTTAAA GCATATCATC ATTTACTTTT TAATCAGAGT 480  
 TACATCCAAA TGATAGATTT CACGTTATAC CTTACGTAT AATATTATGT ATCGTTTGTA 540  
 45 AGCAAATGAC TAAAAGTCTA TTAATATATA CATTTAATTA ATTGAAAGGA TTGACTACAT 600  
 GATACAAGAT GCGTTTGTG CACTTGATTT TGAAACAGCA AATGGTAAAC GTACAAGTAT 660  
 50 TTGTTCTGTC GGAATGGTTA AAGTCATTGA TAGTCAAATA ACAGAAACAT TTCATACTCT 720  
 TGTGAATCCG CAAGACTATT TTTCACAACA AAATATTAAA ATTCATGGCA TACAACCAGA 780

55

	aGATTTACCT GTTGTGCGAC ATAACGCGGC ATTTGATATG AACGTCTTAC ATCAAAGCAT	900
	TCAAAATATT GGTTTACCAA CTCCAAATTT AACTTACTTT TGTAGTTATC AACTTGCTAA	960
5	AAGAACCGTT GATTCGTATC GATACGGTTT AAAACATATG ATGGAGTTT ATCAATTAGA	1020
	TTTTCATGGT CATCATGATG CATTGAATGA TGCCAAAGCA TCGCAATGA TTACTTTTAG	1080
	GCTACTGAAA AATTATGAAA ATTTAACATA TGTAACATA ATTTATGGTA AAAATCTAAA	1140
10	AGATAAAGGC TAGGACTAAA TAAAATACTC CCTTCAAAAG TAAGCATTGT AAAAATGTAA	1200
	ACTTTCAGG GAGCTTTATT TTATATAAAG TCATATATCG TCATATTTTT ATAAGTTGAT	1260
15	TGTTCTAAAT TACCTACAGT GACACCAATA AGTCGAATTG GTACATCAGG GTCTTTTAAA	1320
	TCGTTATAAA GTAAATATGC AATATTATAA ATATCTTCTT CAGAACTAAC CGAATCTCTT	1380
	AAACTCATCT GTTTAGATAG CGTTTCAAAT TGATAAGTTT TAATTTTAAAC CGTTACAGTT	1440
20	TTAGCTGACT TCTGTAATTT ATTTAGACGT TCAGCTGTTT TACCTGnACA ATTCCCATAC	1500
	TTTTCTTAAA ATCTCTTCAT CATCATTCAC GTCTGTTGCA AATGTGCGTT CAGTCCCTAC	1560
	TGATTTTCTT ACTCTTGATG ATTTCACTTC ACTATGGTCA ATACCGCGTG CCTTGTTATA	1620
25	TAAACCCCGA CCTCTTTTTT CAAACAAACG TATTAATTCA AATTCCGTTT TCTCATATAA	1680
	ATCTCTACCG TTAAAAATAC CATTATCATG CATTACTTTT TTGGAAGCTT TACCTACGCC	1740
	TGGaAAATCT CCAATATCCA ATGTCATCAA AATATCATGG aCATTTTGAT AATCAATCAC	1800
30	AGTCATACCA TCAGGTTTAT TCATACCACT CGCTAATTTA GCTAAAAATT TGTTATAAGA	1860
	AACACCTGCA GATGCTGTTA AATGTGTCTG CTCTAGAATA TCTTTTCTAA TATACTGAGC	1920
	AATTTTCGAA GCAGGAAGGT CTGGTCTCAC TAATTCTGTA ATATCTAAAT ACGCTTCATC	1980
35	CAATGACATC GGTTCACCT TATCTGTATA ACTTCGGAAA ATAGACATAA TCTGCGCAGA	2040
	TGTTTCTCGG TAAGCACCAA AATTACTTGT GACAAAGTAT CCATTGAGC ATAATTTATG	2100
40	CGCTGTGAC ATAGGCATTG CTGAATGGAC GCCGTATTTT CGTGCTTCAT AGGATGCCGT	2160
	AGAGACAACA CCCCTACTGC TTGCTTTACC ACCAACAATG ACTGGTTTCC CTTTCAATTT	2220
	GGGTTATCT CTCATTTTCA CTTGTGCAAA AAAATAGTCC ATATCTATAT GAATAATTCG	2280
45	TCTCTCAGTC AAGTGCTCAC CTCCCTACTA ATTTTACTT TTATAACGCA CAAAAATATC	2340
	TCAACATAAT TATACGCTGT GTACGATTTT TTTACATAAA TCTTGCACTT AGCGATAACT	2400
	ATATTGaGAT AACTACAAGT TGTTATaAAA TCAATTGCTA TTTAAGCATG ATGATGAAGA	2460
50	CGATTGAGTA AGAAAACATA GGTAATCTGA AATAATTCAA GCAAATTCAT TTTGTTGGTA	2520
	TCATCATATT AAAATTTATT ATTGAGTCGG CTTTGTATGA TACAAATAAA TACTATCTTC	2580

55



	AAAGCAATAA GCGGTATGCA TACTAAACAT AAAAATAAGT GATGAATAAC CAAATACCTT	2700
	AATTAAAATA AGCAAGCCAG TACTTAATAG GATTAGTGGT GACAGCATAA TAATTGAGAA	2760
5	TTGCCATTG TTGAAGCAAG CATCTGCTGT TTGGAATAAG ATTCTGTCTT TTTTATATT	2820
	AAACATAGGT TTGCTATCTT TTTTAAATAA AAGAAATAAT GCTCTATGGA TAAGTTCATG	2880
	TAAATCAAT AAAATAATGA ATCCAGCAAA CCCATATACA AGATTGATGA TGATATTTTG	2940
10	ATCGACAACC GCTGTGACAC CTAACGCCCA CTTATACGTA AATAAAATCA CGAATAACGC	3000
	AATAACAAGT TGCAAGATAA TAAACCTTCG CATTTGAAAA TTATTGTGCG TTAAATCAAT	3060
	TTTATGCATT ACCAACCCTC CCGATCATGA CATTCTTATT CTTCTTTAAA TATAGTATAC	3120
15	AATGTCACAT TTAATTTAAA AAGTTCATAT CAAGAAAGTA AATTGGCTGT AATAAAATT	3180
	TAATATACGA CTTCTTCTT CACTTATTAA GCGGAAATTT TATCTCAAAT CATGTGCGCT	3240
	ATTTCAAATT GAATAATGCC ACTGTCTCAA CATGTGTGT TTGTGGAAAC ATATCTACCG	3300
20	GTGTTACCTC TTCAAGTTGA TATTTTTTCAG CTAATAATAA TGCATCACGT TGCTGTGTTG	3360
	CGGGATTACA TGAAATATAG ACAATACGCT TAGGTTCTAA TGTAAACAAA GTCTGAATAA	3420
25	ACGTTTCGTC ACAGCCCTTT CTGGCGGAT CAACCATAC AACATCTGGT TTAATCCCTT	3480
	GTGCTTTCCA TTGTAAATA ACTTCTTCAG CTTTCCACA GACAAAAGTT GTATTATTGC	3540
	ATTGGTTTAT AGTCGCATTT TGTGTGCGT CTTCAATTGC AGAAGGTACT ACTCAACAC	3600
30	CGTATACATG TTTTGCAAGT GGTGCCATAT ATAGCCCTAT TGTCCAATA CCACAATAGG	3660
	TATCTAATAC AACTTCATTA CCTGTCAATT GCGCATACTC AATTGCTTTA TTATATAATT	3720
	TCTCTGTTG TTCAGAATTA ATTTGGTAGA ATGACTGATC ACTTATTTTA AATGTACTAT	3780
35	CTGTTAATTG ATCAATAATT GTATCTTTAC CATATAGCGT TATAGATTGA CGTCCATAA	3840
	TAACATTAGA GTGGCTATCA TTAATGTTTT GTTTAATGCT TGTCACATTA GGAAATGCAT	3900
	CTAATATCTT CTCAACAACA GCATTTTTTT GTGGCCACTT TTTACCATTA GTTACAAAAA	3960
40	TAATCATCAT TTCGTCTGTA TGATATCCTG TTCTTACAAC CAAATGTCTC ATTAAACCTT	4020
	TTTTCAATTG TTCTTGATAA ATACTTACAT TTAAATCTT TAAAATAGAT TTAACCTCAT	4080
	TCATCACTTC TTGATGTTGT GAATCTTGTA TTAAACAAC TTCCATGTCA ATAATGTCAT	4140
45	GGCTTCTTTG ACGATAAAAG CCCATAATAA CTTCAATTCTG TTCATTCTTA CCAACTGGAA	4200
	TCTGGGACTT GTTTCGATAT CTCCAAGGAT CTGTCATGCC AACTGTATCG TTAATCTTAG	4260
50	AATTATCAAA ATGCGCTTTT CGCTGAAACA AATTAATCAC TTGTTCTTTT TTCATTTCAA	4320
	GTTGTGCTTC GTATGATAAG TGTGAAGTT GGCACCCACC ACAACGTTCA TAATATATAC	4380
55		

AGTTCCTTTT TACTTTGATA ATTTTATATT CAATTTGTTC ATTAATTAAA GCTTGTGGTA 4500  
 TGAAAATAGG AAAGCGATCT ATTTTACGA CACCATGGCC TTCATGCGTT AAATCAACAA 4560  
 5 CTGTTCCCGT TTTTATGTCA TTTTATAGCTA TTGCTTGCAA AATTTTACCT CCAAAATGAA 4620  
 CAGGTTAGGA ACAAATTAT GCGCTTCCTA ACCTGCCATT ATATATTTCA CTATTTCTGT 4680  
 TTATCTTCG ATTAATTGT CATCAACATG ATCATTATTT ATTAACCTT CATTTACAAT 4740  
 10 ATCATTAGGT GCAAAGACAT CTATATGACG TTCTAGATTT AAGAAATTCG CTGGTAATTT 4800  
 ACCACCATAT TCTCCATCTA CATTAGTTG TAAGTCTGTG AATGATGAAA TATTAATTGC 4860  
 CTTGCTTTT TCATAAATAA CTTTAGGATG CTTAGTATGT TCTCCTCTG AAGCTAAAGT 4920  
 15 CATAATATGA CCAAGTTCTG CAAGGTTTGA TTTTCAACT ATAATTAACG TAAATAGCC 4980  
 GTCATCTAAC TTAGCGTCCG GCACTAATTT TTCAAATCCT GCCATTGAAT TTGTTAAACC 5040  
 20 TAAAAAGAAT AATAATGCTT CTCCTTGGA AACATTACCA TCATATTCAA TTCTTAAATC 5100  
 TACAGCTTTC ATTTGAGGTA ACATTTGGA ACCTTTGATG TAATAAGCAA ATGGACCAAC 5160  
 AATAGATTTC AATTTACTCG GTGTTTCATA AGAGACTTGC GTCAATTGTC CGCCTGCAGC 5220  
 25 TAAATTAATA AAGTATCGAT TATTCATTTT ACCAATATCT ACTTTAGTAG AATGACCTTC 5280  
 AATGATGACA TCAAGTGCCC CCATGATGTC ATTAGGTATA TGCAATGCAC GTCCAAAGTC 5340  
 ATTAACAGTA CCCATAGGAA TGACACCTAG CTTAGGACGA TTAGGCTTTT CTGCGATACC 5400  
 30 ATTAACACT TCATTTAATG TTCCATCACC ACCTGCAGCG ATTAATACAT CATAATTTTC 5460  
 ATGCATAGCT CTTTCTGCTT CAAGTGTGGC ATCACCTATT TTCTCGGTG CATATGCACT 5520  
 CGTTTCATAT CCCGCTTTT CTAATTTTAT TAAGGCATCA GGTAATTCTC TTTTAAATAG 5580  
 35 CTCTTTACCT GATGTCGGGT TATAAATGAT TCTAGCACGT TTCCTCATAT CTTATCCCTC 5640  
 TACTTAAAAT TCATATATTT TAACTTCATC TTTGTTTCGT CTAATAGGGA GTGGGACAGA 5700  
 AATAATATTT AACAAAATTT ATTTCTGTTCT ACCCCAATT GCATTGTCTG TAGAATTTCC 5760  
 40 TTTGAAAATT CTCTATGTTG GGGCCCCACC CCAACTTGCA CATTATTGTA AGCTGACAGA 5820  
 AAGTCAGCTT CTTGTTTGG GGGCCCCGCC AACTTGACA TTATTGTAAG CTGACAGAAA 5880  
 45 ATCAGCTTCT ATGTTGGGGC CCCACTAGAA TTGAAAAAG CTTGTTACAA GCGTATTTTC 5940  
 TTTCAGTCAA CTACAGCCAA TATAACATTG TAGTGCCTAG GACATTGAAT TTATGACCCA 6000  
 GGCTCAGTCT TATTTTCATCA TTCTTAATAT CGTTAAAGAC CAACTTGTAT CTTAAACAAA 6060  
 50 TACTATCTCA ATATGTACAA AGCTTGTTAT TTATTCAGCA TTTTGTGCCG TTCTTCATTA 6120  
 TATAGCTTCG TCAGTTATGC TATTTTACCT TTAATATGAT GTTGTAATA TAATGTTGTC 6180

55

	AACGCATTAA TAAAATTAAT ATTTTACCA TTAACATGTA CAATGAATAA AGTTAAAAGT	6300
	AATTTGACTT CTATAGATAT AAATAAACCC TCGATTGCAT CTAAGTCAGC AATCAAGGGT	6360
5	TTATTTTTTA AATCTTCATA GTTTGATGAT TTAAATTATC TTTTATCTAA TTCTTGTTTT	6420
	AATAGTTGAT TTAATAATTG TGGATTAGCT TGACCTTTAG ACGCTTTCAT AATTTGACCA	6480
	ACTAAGAAGC CCATAGCTTT GCCTTTACCA TTTTGTAAAT CTTCAACTGA TTGTTGTTA	6540
10	TTGTCTAATG CTTCAATTAC AAATTTTAGA AGTGTGCTT CATCAGAAAT TTGAACTAAG	6600
	CCATTATCTT CCATAATCTG TTTAGCATTG CCACCTTTAG CTGCTAACTC TGGGAAGACT	6660
15	TTCTTCGCAA TTTTACTGCT CATTGTTCCG TCTTCGATAA GTTTAATCAT ACCTGCTAAA	6720
	TTTCTGGTG TTAATTTAGT ATCTAATAAT TCTACTTGAT TTTTATTTAA ATATTCGTTT	6780
	ACGCCACCCA TTAACCAGTT AGATGTTAAT TTAACATCTG CACCGTGTTC AATTGTTGAT	6840
20	TCAAAGAAAT CTGACATTTT TTTAGTCAAT GTTAATACGT GTGCATCGTA TGCAGGTAAA	6900
	CCTAATTCAT TTACATACTT AGCTTTACGT TCATCTGGTA ATTCAGGAAT TGTCTGACGA	6960
	ACACGCTCTT TCCAAGCATC ATCAATATAT AAAGGTACAA TGTCAGGCTC TGGGAAGTAA	7020
25	CGGTAATCAT CAGAACCTTC TTTAACACGC ATTAAAATTG TTTTACCTGT AGATTCATCA	7080
	AATCGACGTG TTTCTTGTC GATTTCTCCA CCATTTAACA ATTCTTCTTC TTGGCGTTTT	7140
	TCTTCATATT CTAAACCTTT ACGTACATAG TTAAATGAGT TTAAGTTTTT CAATTCGGCT	7200
30	TTAGTACCAA ATTTTCTTG ACCATATGGA CGTAAAGAGA TGTTAGCATC ACAACGTAAA	7260
	GATCCCTCTT CCATCTTAAC GTCTGATACA CCAGTGATT GAATAATTGA ACGCAATTTT	7320
	TCTAAATATG CATATGCTTC TTTAGGTGAA CGAATATCTG GTTCAGATAC GATTTCAATT	7380
35	AGCGGTGTAC CTTGACGGTT CAAGTCAACT AATGAATACT CACCTTTATG TGTTGACTTA	7440
	CCAGCATCTT CTTCCATGTG AAGACGAGTA ATACCGATTG GTTTTGTTTC ACCGTCGACT	7500
40	TCGATATCGA TATATCCATT TTCACCAATT GGTTGATCAA ATTGAGAAAT TTGATATGCT	7560
	TTTGATTAT CTGGATAGAA ATAGTTCTTA CGGTCAAAC TAGATTCTGT TGCGATTTC	7620
	ATATTTAGTG CCATTGCAGC ACGCATTGCC CAGTCTACTG CACGCTTATT AACAACTGGT	7680
45	AAGACACCTG GATATGCTAA GTCGATAACA TTTGTATTTG AGTTAGGTTC TGCTCCAAAA	7740
	TGCGCTGGTG ATGGAGAAAA CATTTTTGAG TCCGTTTTTA ACTCTACGTG AACTTCAAGT	7800
	CCTATAACTG TTTCAAAATG CATGATTTCC ACTCCTTATA ATTTTTCATA AACGTCATGT	7860
50	AAATTGTATT GTGTTTCATA TTGATAAGCG ACACGATATA ACGTTTTTTC ATCGAATGGT	7920
	TTACCAATGA ACTGTAAACC GATTGGTCGG CCATTTGATT GTCCACAAGG AACAGAAATA	7980
55		

	GGATCATCAA TTTCTTCACC TAAATTAAAC GCaGTgTnAG GCGCTGTTGG ACCAACTACT	8100
	ACATCATAAT TTTCGAATAC TTTATCAAAG TCATTTTTTAA TCAATGTTCT AACTTTTTGA	8160
5	GATTTTTTAT AGTAAGCATC ATAGTAACCT GAACTTAATG CAAATGTACC TAAGAAAATA	8220
	CGACGTTTTA CTTCTTTACC GAAACCTTCA GATCTTGACA TTTTATATAA TTCTTCTAAT	8280
	GAATGAGCTT CTTTAGAATG ATAACCATAA CGAATTCGGT CAAAACGAGA AAGGTTTGAC	8340
10	GAAGCTTCTG ATGATGCAAT CACGTAATAT GATGGAATAC CAAATTTAGT ATTTGGCAAT	8400
	GATACTTCCT CAACGACAGC ACCTAAAGAT TTAAAGTTT CTACAGCGTT TTGAACTGCT	8460
15	TCTTTTACGT CATCAGCTAC ACCTTCACCT AAGTATTCTT TAGGTAATGC AACTTTTAAT	8520
	CCTTTAATAT CTTTACCAAT TTCAGATGTA AAGTCTACAT CATCAACTGG TGCACCTGTA	8580
	GAGTCATTAA CATCTGCACC AGAAATAGCT TCTAATACGA TTGCATTATC TTTTACATTT	8640
20	CGAGTCAATG GACCAATTTG GTCTAATGAA GATGCAAAAG CAACTAATCC AAATCGAGAT	8700
	ACACGACCGT ATGTTGGTTT CATACCGACA ACGCCACAAT ATGCAGCCGG TTGTCTAATT	8760
	GAACCACCTG TGTCTGAACC TAAGCTAAAT GGTACTAAGC CAGCTGCAAC TGCTGCTGCA	8820
25	GATCCACCTG ATGAACCACC TGGCACTGCT TTATGGTCAA ATGGGTAAAC TGTTTTTTTTG	8880
	AAATAAGATG TTTCTGTTGA ACCACCCATT GCAAACCTCAT CCATATTTAA TTTACCGATT	8940
	AAAACGGCAT TTTCATTATG TAGTTTTTCC ATTACAGTAG ATTCGTAAAT TGGCACAAAA	9000
30	CCTTCTAACA TTTTACTTGC ACATGTTGTT TCTAATCCGT TTGTAATAAT GTTATCTTTT	9060
	ATACCCATTG GAATACCAAA TAATTTGCCA TCCATTTGAT CTTTGTCTTG TAATTCATCC	9120
	AATTCTTGCG CTTTTTTGAT TGCATTTTCT TTATCCAGCG CTAGAAAAGA CTTAATTGTT	9180
35	GGATCAGTCT CTTCAATTGC ATCATATATA TCTTTAACAA CATCAGATGG TTTGATTTTT	9240
	TTGTCTTTTA TTAAAGTTAA TAAATTCTCA ACCGATTCTG AGCGAATGCT CATCTTACGC	9300
40	GTCCTCCTCA TTCATGATTG TAGGCACITT AAATTGTCCA TCTTCTGTTT CTTTGGCATT	9360
	TTTCAAAGCT AATTCTTG TG AATACCTTT AATTGCTTTA TCTTCACGTA AAACGTTTTG	9420
	TAAATCTAAA ACGTGATATG TAGGTTCAAC GCCTTCTGTA TCAGCGCTAT CATTTTGTTT	9480
45	TGCAAAATCT AAAATGCTTT CTAATGTGTT GGCCATTCTT TCCGTTTCTT CAGGAGAAAT	9540
	TTGAAGTCTT GCAAGATTCT CGATATGCTC AACTTCTTCA CGTGTACTT TTGTCATTAA	9600
	TAAAGCCTC CTTTAAGTCA TTCATCACTA AATTGTATCA AATTCCAAT TAAAAATCTA	9660
50	AGTATTTATG AGGTGCTACT TTAATTTTCAT ATAACTGTA TAAACATTAT CATTCGTTTA	9720
	TCAATCATT TTTTATGAAA ACAACACTCT TTTAATATTA GACAACCCAA TTCAATATTA	9780

55

TATATTGGTA TGCAAGTATT TCAAAAAGAA TAAATTTAAT TTTCCTACTT TTCTAAACAT 9900  
 TTATCTTTAT GTATAATGTT TTCAAGTAAC TAAATTATAA ATTAAATAAA GGGAGTGTTT 9960  
 5 ATCATGCTTA CAATGGGGAC AGCATTAAAGT CAACAAGTAG ATGCCAATTG GCAAACCTTAT 10020  
 ATTATGATTG CCGTCTACTT CTTGATACTA ATCGTTATTG GCTTTTACGG TTACAAGCAA 10080  
 GCAAACCTGGTA ACCTAAGCGA GTACATGTTA GGTGGACGTA TATTGGACCG TATATTACTG 10140  
 10 CATTATCAGC TGGAGCTTCA GATATGAGTG GATGGATGAT TATGGGGCTA CCTGGTTCTG 10200  
 TCTATAGCAC TGGTCTATCA GCTATGTGGA TTACAATCGG TTTAACATTA GGTGCTTATA 10260  
 TAAATTACTT TGTTGTTGCT CCTAGACTTC GTGTTTATAC CGAATTAGCT GGAGATGCAA 10320  
 15 TTACATTACC AGATTTCTTT AAAAATCGTT TAAACGATAA AAATAATGTG TTAAAGATTA 10380  
 TTTCTGGATT GATTATCGTA GTATTCTTTA CATTATATAC ACATTCTGGT TTCGTATCTG 10440  
 GTGGTAAACT ATTTGAAAGT GCTTTTGGAT TAGATTATCA TTTCCGTTTA ATATTAGTTG 10500  
 20 CTTTCATTGT CATTCTCTAT ACTTCTTTG GTGGATATTT AGCTGTATCA ATTACAGATT 10560  
 TCTTCCAAGG TGTCATTATG TTAATTGCGA TGGTTATGGT CCCTATTGTT GCTATGATGA 10620  
 25 ATTTAAACGG CTGGGGAACG TTTTCATGATG TAGCAGCTAT GAAACCTACA AATTTAAATT 10680  
 TATTTAAAGG GTTATCATT ATAGGAATTA TCTCTCTATT TTCATGGGGA TTAGGTTATT 10740  
 TCGGTCAACC TCATATCATT GTAAGGTTTA TGTCTATTAA ATCACACAAG ATGCTACCTA 10800  
 30 AAGCTAGACG TTTAGGTATT AGCTGGATGG CTGTTGGTTT ATTAGGCGCT GTGGCTGTTG 10860  
 GTTTAACAGG TATTGCATTC GTACCTGCTT ATCATATTAA ACTAGAAGAT CCTGAGACAT 10920  
 TATTCATCGT GATGAGTCAA GTACTCTTCC ATCCTCTTGT AGGTGGTTTC TTAATTGCTG 10980  
 35 CGATTCTAGC TGCAATTATG AGCACGATTT CTTCACAATT ACTTGTAACA TCTAGTTCAC 11040  
 TAAAGGAAGA CTTTTATAAA TTAATTCGTG GTGAAGAAAA AGCTAAAACG CACCAAAAAG 11100  
 AATTTGTTAT GATTGGAAGA TTATCTGTAT TAGTTGTAGC AATTGTTGCC ATCGCGATTG 11160  
 40 CATGGAATCC AAACGACACA ATTCTAAACT TAGTAGGTAA CGCTTGGGCC GGATTTGGTG 11220  
 CATCGTTCAG TCCACTTGTG CTATTGTCAC TTTACTGGAA AGGTTTGACA CGTGCCGGTG 11280  
 CTGTAAGTGG AATGGTTTCA GGTGCCTTAG TCGTTATCGT TTGGATTGCA TGGATTAAAC 11340  
 45 CATTGGCACA TATCAACGAA ATATTCGGCT TATATGAAAT TATTCCTGGA TTTATTGTAA 11400  
 GTGTAATCGT TACATATGTT GTAAGTAAAC TTAATAAAAA ACCTGGTGCA TTTGTTGAAA 11460  
 50 CTGACTTAAA CAAAGTTCGT GACATCGTTA GAGAAAAATA ATTCATAAGT CTTAACAAAT 11520  
 TAAAAAGGTA CTAATGTTAA TCAAAATTAT GACTAACATT GGTACCTTTT TATTATCTTT 11580

55

	AATTAAAGCA CGTGGTTGGT TACCATCTTT AATACGAATT TCATAGTTAT CGATTTTATC	11700
	GAAATATTTA TTCGCTTGTT CAGTAACGTA CTGTGTAATA CCAATTGTTT CAGCTTGTCC	11760
5	ATAGTAATCG ATTGGTAAAT CTACTIONTAA TCGTTGTGGC TTTTATCAA CAAATTTAAC	11820
	TTCCCTACT GCTTGTGTGA AATTAGAAAA ATATGATTGC AAATTATCAT TAAATTGCTT	11880
	GAAATTATTA TTTAAATTTT CATCATAATC TGCTGCTGTT GAAGAAGGTA ATAAAGCTGA	11940
10	TTTTTCATTG ATATTATGCC ATTCAATTAAG CTTTGTGTTGA CTCTTTTCTG CAGTCGCTTG	12000
	AGTGATAAAT TCACCTGGTG TGATTGAATC TTCACTTGAT TGCTTATAAA TTGCAAAATG	12060
	AATTGGTATA TCTTTTAAAT CATCATTTTC ACGTAACCTT GATAATATCT CACTAGCCAT	12120
15	TTGTTTACCT TGCTTTTTTAA CTcGCTATCA TCTAGTTTTT TACTAAAAGT CGATCCATCT	12180
	TTTTCTTTTT TATAGTAATA AACACTATTC ATAGCTAAAC CAATCGTCAT ACCTTTAATA	12240
20	TTCTTACCTT TTGTATCTCC ACCACCATAA AAATCTTGCT CTAAAATGTT AGATAAATAG	12300
	GCTGGTGATT TTTCTGCAAT CTTTTCAGGA TCTGTTTCAC CTcCGTGTGA TGGATTAGT	12360
	CCTAAATTTT CATTCGCTTT CTGTCTTTT TTATCTTTTT CAGACATTTT ATCGATTTC	12420
25	CGTTTTGTAT ACTTAGGATT TAAATAGGCA TTAATTGTTT TCTTGTCCAA AAATTGACCA	12480
	TCTTGATACA AATATTTATC TGTGGGAAAT ACTTCTTTAC TTAAGTTCAA TAAACCATCT	12540
	TCAAAGTCGC CGCCATTATA ACTATTTGCC ATGTTATCTT GTAAAAGTCC TCTTGCCCTGG	12600
30	CTTCTTTTAA ATGGTAACAA TGTACGATAG TTATCACCTT GTACATTTTT ATCCGTTGCA	12660
	ATTTCTTTTA CTTGATTGA ACTATGTGA TGTTTTTGAT TATCTTTTCC AGCCTGGTCA	12720
	TCCTTATGGT TACCACAAGC AGCGAGTATA AAGATAGCTG TAATCAATAA TACTAATGTA	12780
35	CGCTTCATCG ACATACCCCT CTAACATTTT AATTCATTTT GCTTATCTAC AAATTGTTGC	12840
	TCTGTCCAAA TTTCAATACC TAAACTTTGT GCTTTTGTTA ATTTTGAACC TGCATCTTCA	12900
	CCAGCAATAA CGACATCTGT ATTTTATAGTA ACGCTACTTG TAACTTTAGC ACCTTGAT	12960
40	GCAAGCCATT TAGATGCTTC ATTGCGTGTC ATTTGATGTA GCTTACCAGT CAGTACTATC	13020
	GTTTTACCAC TAAATTCAGG ATGTCCTTCA ATATCTGATG TTTTGATACC TTTATAAATC	13080
45	ATATTAACAT GTTTATCTTT TAATTTTGA ATTAAAGCAC GAATATCTTC ATTTTCTAAA	13140
	TAAGTAACTA CAGATTGTGC TACTTTATCA CCTATATCAT GAATTTCTAC TAATCCGCT	13200
	TCAGTTACCG TTAGTAATCG ATCTATCGTT TCATATTTTT CTGCTAACAC TTGGCTCGCT	13260
50	TTAACACCTA AATGCCTAAT ACCTAGACCA AATAATAAAT TTTCTAAAGA GTTGTCTTAA	13320
	GCTTGTTGAA TGGCAGCTAA TAAATTATCA ACTTTTTTCT GCCCCATTCT GTCTAAAGGT	13380

55

TAAAGCTGTT GAATAATTTT AGTGCCTAAA CCATCAATAT TcATGGCTTG TCTTGaTACA 13500  
 AAGTGnATCa ATCctTcAAC AAGTTGTGCT TGGTCATTTT GG 13542

5

(2) INFORMATION FOR SEQ ID NO: 155:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1893 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

15

CAGTAAACAC CTCTGATTAC GAATATTTAT ACATTTATTT TAACACATGC ACTGATTTAC	60
GACTACTAAA CACCTTTTACG TAAAAAGGGT AAACATGGTT TATCTATCTT GGTTATCTAT	120
TTATAAATAT TThTCATATT ACGCATAACA ATTGCTTAAA ATATGTATAA AAATGAATAT	180
ATGTGTAATA AACTTGCTAA TTATTAGATT TAATAAGCGT CAATTGTTTG AACATATTtA	240
ATTAAATCA CATTGATATC ACAGATACGA ATATTGTCGT ATAGAAATTG AAAATTCTAT	300
TTTTTAAATG AAAGTCTTCA ACATAATTTT AAGTTTCAAC ATGAGAAAAA TCGATTAACA	360
AACAACGTCa GTTGAATATG CCTTTTGAGA CATTTCAAAC TTTACAATTG TTGCTAATCG	420
ATATATTTGC TTTTAGTGAT CCCTGCTATA AAATAAATCA ACGATTTCTA ATAAGTGTtT	480
TGTATTGAAT TGTTCACTAA TTTGCGTTAG TTCATCCACT GCTGCGTCTC TATGATAAGT	540
CAATTTATCT TCTGCGCCAT CTTTCCCTAA TAAACTCAGG TACGTACTTT TATTATTTTC	600
AAGATCGCTG CCCACTTTTT TACCTAACTT TGCTTCATCA CCATAGCAGT CTAATAAATC	660
ATCTTTAATC TGGAACATCA TACCTAAATG ATAACTATAA CTTTCTAAAT GTTCTTTAGT	720
TGTATCATCG ACATTAGCGA TATCTGCTGC ACTCATAACC GCAAAAGTTA ATAATGCTCC	780
TGTTTTTGTt TTGTGTATCA TTTCCAAAGT TTCAAGATCA ATTGGTTGGC CTTGCTTTG	840
CATATCTAAC ATTTGACCGC CGACCATTC AACATGACCA CTTGCTATTG ACAGCCGTTG	900
TAGAACTTTT ATTTTACTT CATCAGTTAA TCTATCATCA CTTGAAATAA GTTCAAATGC	960
TTTAGTTAAT AAAGCATCAC CTGCTAATAT CGCAGTCCAC TCACCATATA CTTTATGATT	1020
TGTTAATTTT CCTCGTCGAT AATCATCATT ATCCATCGCT GGTAGGTCAT CATGAATAAG	1080
TGAATATGTA TGAATCATTT CTAGTGCAAT TGCCTCTTC ATACCTAACT CATACTCGGT	1140
ATTTAGTGAA TCTAAAGTGA GTAATAACAG AACTGGTCGG ATGCGTTTAC CTCCAGCATT	1200
TAATGAATAC AACATACTTT CTTCTAGCTG AGTATCCATT ACTGATTTAT TTATCGCAAC	1260

55

CATCCTCAGC TTCTTCITTT ATTAAGTCAT TCACCTTTTT TTCGGCATT TTTAAAGTTG 1380  
 TGTCACAAGC TGCTGATAGT TTCATACCAC GTTGATATAA ATCTAATGAT TCCTCTAAAG 1440  
 5 ATACTGTTTC ATTATCTAAT TTTTGAACAA TTTGCTCTAA TTCTTGCATC ATTTCTTCAA 1500  
 AACTTTGCGT TTCTTTAGTC ATTATTACAC CTTACTTTCG TAACTTTTCG ATCTACTAAG 1560  
 CCATCTTTCA TTGTTAACGT CAATTGATCA TTTTCTGTAA AATCTTTAGT ACTCGTAATG 1620  
 10 ACTTCGTCIT TTTTATTAAC AATTGCATAT CCACGCAACA TTGTATTAGT TGGACTTAAA 1680  
 TTGTTTAAAGT TTTCTACTTT ATTTTTCAAA TCATTTTAT AACTTAATAT CTTAGAATTC 1740  
 AATAATTTAA CAAGTTGGTT TGTCAATTGA AGATTATnTT GTTGTCTTG ATTAACACTA 1800  
 15 CTTAGTAATG CTTTTAAATn ATAACGTTGG TGCAACAGCA TTAAATCGAG GCCCCGGTGG 1860  
 TCCAAAGTTG CCCGAATTnG TGGTTTCAGG CCC 1893

20 (2) INFORMATION FOR SEQ ID NO: 156:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 821 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:  
 30 AAAATATATT CCTTCACTTA ATATTCAATT AGAGAAAAAC ATGGTGATTG TAATATGTTG 60  
 TGCAATATTT CTGGGTGTTT TAATACTTTT TTTATTTCTG AATCGTAAGC TAAGGTTGGA 120  
 AATTTATAAT AATAACTCTA GTAAAGGGAA AATAATTTTA TTTCTTCAT TAAAAAAGTT 180  
 35 TTGTTTCACA ATATTTTATT ATTTTTTATT TGGCGGTCTT TCAATAATGG CTCTAAGTAT 240  
 GTTATTAACT TTAAATCCTC AAAATATAAT AGGCTTTTATT GGTTGGTTGG TAATGACTGC 300  
 AGGTTTCTTT CTGTTAAACA TGTCATCGAT TATTGACAAA AAAATTTATG TATTATCTAA 360  
 40 AACTAACACG GTGGA AAAAT GATGGTTTAG CTGGATTTAC TGCAGGTTCT ATTTCCGGCA 420  
 TACTTGTATA TTGGACCAAT CAAAAAATG AATTTGGAAT AAAAGATAAA AACGATTGGA 480  
 TAGGACATAA ACTAGACGTT GGTATAGATG CTGTAGAAAA ATCTGCAGAA AAAACAGTAG 540  
 45 ATGGTGTTGA AAATGTCATG GTGAAGCTTC AAAAGTATT TCTAATCATA TAAGCCCTAA 600  
 GAAATGGAGC TGGTAAATGT TGCTATGCGA ATCTAAAATC ATCAATAAAA ACCCAAATA 660  
 50 TAGAATTATT AAATATAATG ATGAATACTT AATGGTCGAT ATAATAAGCA CTTGGATTAG 720  
 TTTATTTTTT CCTTTTATTA ATTGGTTCAT CCCaAAGaA TACGTCAAAA TTAGTAGAGA 780

55



## (2) INFORMATION FOR SEQ ID NO: 157:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2343 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

```

AGTAAGATAA TTTTCAATTA GAAAATATCT TACTGCTGTT CTCTATTTAT ACAATACTTC      60
GTATTGAATG GcTTTCGCTTT CCTAGGGTGC CGTCTCAGCC TTGGTCTTCG ACTGGCACTG      120
CTCCCTCAGG AGTCTCGCCA TTAATACTAC GTATTAAACAT GTAATTTTAC TTTGAAATAC      180
TTTTAAAAAA TAAGACACTT TGCCCAACTT GCACATAAAT GTAAAATTCA ATAAAATGAA      240
TTTTCTGTGT TGGGTCCCTT CTTATAATTT AATAAATACC ACTAACTAA ATTAACGAGG      300
TGCCTTATGT ATAAAATTA TAACATGCCC CAACTACACT ACCAATAGAA ACTTCTGTTA      360
GAATCCCTCA AAATGATATT TCACGATATG TTAATGAAAT TGTGAAACr ATACCTGATA      420
GCGAATTGGA TGAATTCAGA CATCATCGTG GCGCAACATC CTATCATCCA AAAATGATGT      480
TAAAAATCAT CTTATATGCA TATACTCAAT CTGTTTAAAT ATGTTCAAAG CATTAAAGGTA      540
ACAAGACAAT ATCTAAGATA TCAAAGATAG AAATTTTTTG ACGTTGTTGC TGATTGTAA      600
CATAACCATC AATTTCATAA TTAATAGCAT CAATACGATA AATGGTTAAG CGTACTGAAT      660
CTACAAAGCC ATTATTATAA AATTTAACTT CTACAGGTTG GGCATATTGT AGCGCCTCGT      720
GTAGCCGAAT GTTTAGCTCA GCCAATTGAT CATCTGATAA TACAGGACGT GTAATTTTGT      780
TTTGGTCGAT AATGTATTGT TGAATCGTTT CGAATTGTTT GGGTAATGTT GCAAAAGGAG      840
CCCATTTAAT CATGCCTCTT CCCATAGGTA TATTGTTATC TAGTAATTCT CTTGGAACGT      900
TACGATAATC AGTTTCTTCT TCATAACTTG TCATCCTTAA TTCACCCCAA TCTGATAATT      960
ACATTATACG AACATGTGTT CTATTTTGCA ACAAATTTT TGTGGaAGCA TAAACGCGTT      1020
AATAATTAAT GCTCGTGLAA GTAAAAAGA GGGATTAATT AAAATCGAAT AATGaCATAT      1080
CACaGCAAAT AGTTCTTTTA AAGTAGTTAA ATAGTTTATG CTTTAAGGAA aTGATAaTG      1140
ATTGTWAATT CTAGCTAAAA TTTAATAAAA TGAAAATAAG ACTAACATGG AGGGGTAAAA      1200
GTAATGACAA ATGGATATAT TGGTCTTAC ACTAAAAGA ATGGTAAAGG GATTTATCGT      1260
TTTGAATTAA ACGAAATCA GTCACGTATT GATTTATTAG AAACAGGATT TGAATTAGAA      1320
GCGTCTACAT ATTTGGTGCG TAATAATGAA GTTTTATATG GAATCAACAA AGAAGGAGAA      1380

```

TGT TTTGTCTT CAAAAGCTGG TACAGGTTGT TATGTATCGA TTTCAGAAGA TAAACGATAT 1500  
 TTATTTGAAG CGGTATATGG TGCTGGCATC ATACGTATGT ATGAATTAAA TACGCACACA 1560  
 5 GGTGAAATTA TACGTCTAAT TCAAGAACTT GCACATGATT TTCCAACAGG TACACATGAA 1620  
 AGACAAGATC ATCCACACGC ACATTATATT AATCAAATC CAGATGGTAA GTACGTTGCA 1680  
 GTAACAGATT TAGGTGCTGA TCGTATCGTT ACTTATAAAT TTGATGACAA CGGGTTTGAA 1740  
 10 TTTTATAAAG AATCTTTATT TAAAGATAGT GATGGGACAA GACATATTGA ATTTTCATGAT 1800  
 AATGGAAAAT TTGCTTATGT CGTACACGAA TTATCAAATA CTGTGAGTGT TGCAGAATAT 1860  
 AATGACGGTA AATTTGAAGA GCTCGAGCGT CATTTAACAA TTCCTGAAAA CTTTGATGGA 1920  
 15 GATACTAAAC TTGcAGCAGT GCGTTTATCT CATGaTCAAC AATTCTTATA TGTATCTAAT 1980  
 AGAGGGCATG ATAGCATTGC AATTTTTTAA GTTCTTGATA ATGGTCAACA CTTAGAACTA 2040  
 20 GTAACAaTTA CTGAAaGTGG TGGTCAATTC CCAAGAGATT TTAATATTGC CTCATCAGAT 2100  
 GACCyTTTAG TTTgTGCTCA kGaGCaAGGA GATTcAGTTG TAACTGTTTT CGAAAGAAAT 2160  
 AAAGAAACAG GTAAAATTAC GCTATGTGAT AACACTCGTG TAGCATCTGA AGGTGTATGT 2220  
 25 GTCATATTTT AATCTTTAAT TAATCATGAT AAAAAGAAAA CCATGTTTCC AAAAAATTTG 2280  
 TGTATACCTT GAAATTTATT GnTTTCCAGn ACATCAATTA TGGGAAGCAT GGnTTATTTT 2340  
 TGT 2343

30 (2) INFORMATION FOR SEQ ID NO: 158:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4837 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

AAATTGCCAG TTGGTATCGC TTCTGGTGCA GTAGTCGAAG GTTTCTTCCA AGGTATCATT 60  
 CCGATTGGCT ATATCGTTAT GATGGCAGTA TTGTTATACA AAATTACTGT TGAATCTGGA 120  
 45 CAATTTTTAA CAATTCAAGA TAGTATTACA AATATTTTAC AAGACCAACG TATTCAAGTT 180  
 TTACTTATTG GATTTGCATT CAACGCATTT TTAGAAGGTG CAGCAGGATT TGGTGTACCA 240  
 ATTGCAATTT GTGCACTTTT ATTAACACAA TTAGGATTTA ATCCATTAAA AGCTGCGATG 300  
 50 TTATGTTTAG TCGCAAATGC AGCGTCTGGT GCTTTTGGTG CGATTGGTAT CCCTGTAGGT 360  
 GTTGTAGAAA CGTTGAAATT ACCTGGAGAT GTTTCAGTAT TAGGTGTTTC TCAATCAGCA 420

55

	GGTTTTAGAG GTGTAAAGA AACATTACCA GCAATTTTAG TAGTTTCAAT CACTTATACA	540
	CTTACTCAAG GATTATTAAC TGTATTCAGT GGACCTGAAT TAGCAGATAT TATTCCACCG	600
5	TTATTAACAA TGTTAGCATT AGCAGTATTT TCTAAAAAAT TCCAACCAAA ACACATTTAT	660
	CGTGTTAATA AAGATGAAGA AATTGAACCT GCAAAAGCAC ATTCTGCAAA AGCAGTATTA	720
	CATGCATGGA GCCCATTCAT TGTATTAACA GTCATTGTAA TGATTGGAG TGCGCCATTC	780
10	TTTAAAACT TATTCTTACC AAATGGTGCT TTATCATCAT TAGTATTTAA ATTCAACTTA	840
	CCTGGaACAA TCAGCGAAGT TACGCATAAA CCATTAGTAT TGACTTTAA TATTATTGGA	900
	CAAACAGGTA CAGCTATTTT ATTAACATTT ATTATTACAA TTTTAATGTC TAAAAAGGTT	960
15	AACTTTAAAG ATGCAGGTAG ATTATTCGGC GTTACATTTA AAGAGTTGTG GTTACCAGTT	1020
	CTTACAATTT GTTTCATCTT AGCAATTTCT AAAATCACAA CTTATGGTGG TTTAAGTGCA	1080
20	GCAATGGGTC AAGGTATTGC TAAAGCAGGT AATGTCTTCC CAGTTCTATC ACCAATTTTA	1140
	GGTTGGATAG GTGTGTTTAT GACAGGATCA GTTGTAATA ACAACTCATT ATTTGCACCA	1200
	ATTCAAGCTT CTGTTGCACA ACAAATTGGA ACAAGTGGTT CACTTCTTGT ATCTGCTAAT	1260
25	ACAGTTGGTG GTGTAGCGGC AAAATTGATT TCACCACAAT CAATTGCAAT TGCAACTGCA	1320
	GCAGTAAAC AAGTTGGTAA GGAATCAGAA TTATTAAAA TGACATTGAA ATACAGTGTA	1380
	TGTTTACTAA TATTCATCTG TATTTGGACT TTCATCTTGT CATTATTATA AAAAAACGTA	1440
30	TTTCAAAATA TAAATATACA GAAGGTGAGA TGTTTTCTAA CATCTCATCT TTTTTTATG	1500
	GATCATTAAT GAAAGAAGTT TGACATTATA ATAATGGTAG CGCTTTATGT TAAATGAAT	1560
	AGTGAGTAAT CAGCAATCAA ATTAAATTGG TTGATAGCTG TTAAGGTTTG TGGTTTTGTC	1620
35	TTTGTGCTAT CGCnCATAAA GTATATAATT AAAGTAGTTT CGTTATTATA AAATATTAAT	1680
	ATACATAGTA GATAGTAATA GAGCATCACC ATGGGAACCT ATTGAGACAC TTATTGATTT	1740
	AAAGTGGTAT TAATATGTCG TATTTCTCGA ACGTTCCATT ATTCATTTTA AAAAGGGGGA	1800
40	CTGTATTTGT TATGACAACA CAACATAGCA AAACAGATGT CATCTTAATT GGTGGCGGTA	1860
	TTATGAGTGC aCATTAGGAA CTTACTTAA AGAATTATCA CCTGAGAAAA ATATTAAAGT	1920
45	GTTTGAAAAA TTAGCACAAC CTGGCGAAGA GAGTTCAAAT GTATGGAATA ATGCCGGTAC	1980
	AGGGCATTCA GCACTTTGCG AGTTGAACTA TACAAAAGAA GGTAAGGATG GCACAGTTGA	2040
	TTGTAGTAAA GCAATTAAGA TAAATGAGCA GTACCAAAT TCAAAACAGT TTTGGGCATA	2100
50	TTTAGTTAAA ACAGGACAAT TAGATAACCC AGATCGCTTT ATTCAAGCGG TGCCACACAT	2160
	GAGTTTTGTC ATTGGCGAAG ATAATGTAGC TTTTATAAAA AGTCGTGTTG CAACGTTAAA	2220

55

	GGTACCGTTA ATGATTGAAG GTCGTAAGTC TGATGAACCA ATTGCTTTAA CTTATGATGA	2340
	AAC TGGTACa gATGTAACT TTGGTGCGTT AACTGCAAAG TTATTTGATA ATTTAGAGCA	2400
5	ACGTGGTGTG GGAATTCAAT ATAAGCAGAA TGTATTAGAC ATCAAGAAAC AGAAATCTGG	2460
	GGTATGGCTA GTTAAAGTTA AAGATTTAGA AACTAATGAA ACGACAACAT ATGAATCTGA	2520
	TTTTGTATTT ATTGGTGCTG GCGGTGCGAG TTTACCATTa CTCCAAAAGA CTGGGATTAA	2580
10	ACAATCAAAA CATATTGGTG GTTTCCCGGT AAGTGGATTA TTCCTGCGCT GTACAAATCA	2640
	AGAAGTGATT GATCGTCATC ATGCTAAAGT GTACGGAAAA GCAGCAGTGG GTGCGCCACC	2700
15	AATGTCAGTG CCGCACTTAG ATACACGTTT TG TAGACGGC AAGCGTTCAT TGTTATTTGG	2760
	TCCATTTGCA GGTCTCTCAC CTAAATTTTT AAAACAGGT TCACATATGG ATTTAATTAA	2820
	ATCGGTTAAA CCAAATAATA TCGTGACGAT GTTATCTGCA GGTATCAAAG AAATGAGTCT	2880
20	TACGAAGTAT TTAGTGTAC AATTGATGTT ATCTAATGAT GAGCGTATGG ATGATTTAAG	2940
	AGTCTTTTTT CCAAATGCTA AAAATGAAGA TTGGGAAGTG ATTACAGCAG GGCAACGTGT	3000
	CCAAGTAATC AAGGATACTG AGGATTCTAA AGGTAACTTA CAATTTGGTA CTGAAGTTAT	3060
25	TACGTCAGAT GATGGCACAT TAGCTGCATT ACTTGGTGCA TCACCTGGTG CGTCAACAGC	3120
	TGTAGATATT ATGTTTGATG TTTTACAGAG ATGCTATCGT GATGAATTCA AAGGATGGGA	3180
	ACCAAAGATT AAAGAAATGG TGCCGTCATT TGGTTATCGC tTAACAGATC ATGAGGATTT	3240
30	ATATCATAAA ATTAATGAAG AAGTAACTAA GTATTTACAA GTTAAATAAT AAACGAAACG	3300
	GTAATGTCTT TTTTAATGTG ATAGACATTa CCGTTTTTTA GTGGTTAATA AAAATCATTT	3360
	TAATTGTTTC AGTTGCTTGT TAATAGTGTC TACGTAGTTC TTGTTTTTAA AGAATTGAAT	3420
35	TATCCAAATT AATACATAAA CCACAATGAA GATAATTGTG AATATGATTA GATAATGCAC	3480
	TGTTAGTGGA AACCAACCGG CAAGCATTGC TAAAGGCAAG AATCCGACAT ACGTTGTTAT	3540
40	GAAATGCATT ATAGTTGCTT TAGTAATGCT CCAATCTGTG TATTTAAAGA TAAAATCTCC	3600
	AAGGAAAAAG ACGACGCCTA TGAGTAACCA TAAATGATA GAAATCAACA TTACGGTAGT	3660
	TTCTGTGAAA TCGGTATAAT ACAATATGCC AATAGTTGAT TGTGGGTTCA GTGGATAATA	3720
45	TTTGCCGTCT GCAAATAACA TACTAAAGAA CAGTGAAAGG GACAAACCAA TGATTAAGCT	3780
	AATAAATAAT GAGTTTTTCA AATTTTTCAT ATTGATAAGC GCTCCTTTAT AGATTTTAAA	3840
	TAACGTCTAG AAGAATAGGT GTAGTGTGCA TCTTTAAGAT ACATACGTAT AAGTCCATTT	3900
50	GGCTCTAATA ATAATTTTTT AATGTAATAC TTGTTGACGA TTTCTGATTT GGAAATGCGA	3960
	ATGAAATGTT GTGGTAACTG TTTTCTAGT TCATAAAGTC GTAATTTTAT TTTGAATTTT	4020

55

ACATTAATGA TATGGATTTC TTTGTCTATG TATCCGACTA ATGTATGTGA TTTGTCTAAA 4140  
 TCATTGACTG CATTATAAT ACTTTGAACG TTATCATTCA TTTTAGGTGC ATGTATATCA 4200  
 5 ATATAAGATT CCGTCTCATT TGCATTGATA AATAAATTGA GTTTCATCAT AGGTTAATGC 4260  
 CTCCTTCAAA ATTATTAAAC CATAAATGAC CATCGATATA TTAAATTTT GTTGAATGGT 4320  
 AGAAATTAAA TGTTAAGTGG CTAGAAAGCG CTAATCAATA TAAAAGATAC CTCCTGAAAT 4380  
 10 AAAACAGAA ATGTTTTTTC AGGAGGTAGA GATTAAAGTG AATTATTTGG CAGTGTAATA 4440  
 GTAAAGGTGG TTACATACTC GTTACTTTGT GTGAATTGGA TTGTACCATG ATGCAATTCA 4500  
 ATGATGGATT TTGTAATTGC AAGACCTAAA CCATTGCTAT TATCATGTTT GCTCACTTTA 4560  
 15 TAAAAACGTT CAAATAAACG TGCTTCAGCT TGTGGACTAA TTGGTGAACC ATCATTACTT 4620  
 ATTGTGAAAA TGATATTGTT GTGACTATGT TGCAAAGCGA TGCAATGGC ACCACCAACA 4680  
 TCTGTATACT TAATAGCATT TATTAATAAA TTAATCAATG CTGATGTAA CAAACGTTGA 4740  
 20 TTTCTAGGA AATTGATGAT TCTAGGTCAG CTAATATGAT TAACGACTTT TCATCAGCAG 4800  
 CAAATTGTTT ATGTCGAATG ATATCTTAA TGAGCTG 4837  
 25 (2) INFORMATION FOR SEQ ID NO: 159:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1600 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:  
 35 ACAATTATTG GATTATTATC AAGCAACGTT AATGGATGAC TTCCAATTAC AACAGAAATG 60  
 CCCATAGATT CTAAATCTCT TGCATGAGCA TCTTGTGATA AGTCTTTTCC ATCATTGACA 120  
 40 GTTACATTCG CACCTAATTT ACTTAATAAT TTAGCTGCTT CATAACCACT TTTTGCCAAA 180  
 CCGACAACCTA ATACATTTTT ATTTTCTAAC CCTGTATAAT TAAGCATCTT AATGCACTCC 240  
 AATCCATAAA CCGATTAAAC CTGAAATCAG ACCAACAGCC CAAAATACTG TAACTACTTT 300  
 45 CCATTCGCTC CATCCTATCA ATTCAAATG ATGATGAATC GGAATCATT TAAATATACG 360  
 CTTTCCAGTC AATTAAAGC TAGCGACTTG TAACATAACA GATAATGTTT CAATTACGAA 420  
 TACTAAACCT ATAAAAATTA ATGATAATTC CTGATTAAGC ATGATTGAAA TGGTAGCAAA 480  
 50 TATACCACCT AAAGCTAAGC TACCTGTATC TCCATAAAC ACTTTAGCAG GGTTAATGTT 540  
 ATATGGTAAA AATCCTAAAA GTGCAAACAA CATAATGATA CAGAAAATAC CAATTGCCGT 600

55

TGCTAATCCA TCTAAACCAT CTGTAAATT TACTGCATTA GAAAAACCTA CTTGCCAAAA 720  
 AACAAATGAAA ATAACATATG CAAATGATAG TGGGATTGCT ACATTCGTAA ATGGAATATG 780  
 5 TATGCTCGTA GAAAAATTCA CCAAATGAAA CACATTACTT AAAACAAAGA ATATAATCGC 840  
 AATACCAATT TGCGCCAAAA ACTTCTGTTT ACTTGTTAAA CCTTGTTTAT TCTTTTAAAC 900  
 AACAAATAA TAATCATCTA TAAAACCAAT TAACCCAAAA CCAATCGTCA CAAATAATAA 960  
 10 CAGTATGATT GGATTAGCTT GATCTACAAA TATAATAGCC ACCAAAGACG TTATCACAAT 1020  
 ACTTAATAGA AATGTTAGTC CACCCATCGT TGGTGTACCA GTCTTCTTCA TATGGCTTTG 1080  
 TGGACCTTCT TCTCGAATAC TTTGACCAAA TTTCATCCTT TTTAATGTAG GTATTAAAAAC 1140  
 15 AGGTACCAAA ACAAATGTAA TCACTAGCGC TAATAACGCA TATACAAAA TCATAACTAT 1200  
 CTCCTCTTCT TAATCCAGAC TTTTAAAC ACTAATATAT TATCAATTTT TCAATTAAAT 1260  
 20 AAACAAAGTT GTAATCAAAA TTTATAATTT TTCTTTTCTA CGGCATAAGA GGCCAGTATA 1320  
 AAAAGTTTGC CTATAACAAA CAAGTTAATC TGACCTCGTC TACCTTAAAA TTCTCTATCA 1380  
 ACACTTATTT ATAAAGATTA AATGAAGATG TTGTTTTCTA TCACAGCATT ACTTTAGTAA 1440  
 25 AAACAAATAG TGACAATACA TCCTAATTTA ATGTAGCCAT TCTTGTTAGT CCGACTTATC 1500  
 CTTGTCAGTT TTAATGTCAG ATTTCTCTT ATCATCTGAA TTTGAATCAG AATTATTCGT 1560  
 CGAATTGCTG TCTACATTCT CTGGATGGAA AATTCTACGT 1600

(2) INFORMATION FOR SEQ ID NO: 160:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1186 base pairs
  - (B) TYPE: nucleic acid
  - 35 (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

40 ATTGCCTTTG TTTTAATTTT AAATCAAAAT mGCCTATGAA AGATTTAAAT CAATTAATTT 60  
 CTATAATATT ATCATTTTAA AAGCATATCA TTGTTTAGTT TTTTATAAT TGGATAAATA 120  
 45 CTAATAGTTA CTTTATAAAA CATTACATAG AGAAAGGTTA AGGAGTGAC ATGTCGAAAA 180  
 AGGATCACTC TTCTTCAAAA TACCTTAATT CTGTTAAGGA AGCGCAAGAG GAGTCAAAAA 240  
 AGAAAAATRA AAGTAATCCC AAAATTGATG TTGATCGTAC ATATATTGAA CCTCAACAAT 300  
 50 TCCAATCTAA GAAACCTAAA AAAGATGATC AGGTTTCTT CTTATCAAGA TTAAATAAAC 360  
 CTGCAAAATA TAAGAAAGAC TCTAATTCT TATCATATCT CATCTATCGC ATAGGAAAAG 420

55

TGTTCCTTTT CCTATTAAAC TTATTACCAT TTTTCAATAT TAAGCAGAGT CAAATTACTA 540  
 ATATGTTAAG CAATGCACCC GCTGAAACAT CTA CTACTCTAAT TAAGAGTGTA ATTGGTGATA 600  
 5 TAACTCAAAA CTCCAGTGGT GGCTTATTAT CTATCGGTTT GATTTTAGCA ATTTGGTCAG 660  
 CTTCAAATGG AATGACTGCA ATTATGAATT CTTTCAATGT TGCTTACGAT GTAGAAGATA 720  
 GCCGTAATGG AATCGTATTA AAAC TACTAA GTGTTGCTCT CACTGTAGTT ATGGGCGTTG 780  
 10 TGTTTGTAGT TGCTCTAGCA TTACCAACGC TTGGTTCTGT AATTAGTCAT TTCCTATTCG 840  
 GTCCACTTGG aTTTGACGAA CAAGTGAAAT GGATTTTAA CCTTATTAGA ATTGTGTTAC 900  
 CAATCATTAT TATATTTATC ATATTTATCG TGTATATTC GGTGACCT AACGTTAAAA 960  
 15 CGAAGCTTAA GTCAGTATTA CCAGGTGCAG TATTTACTTC AATTATTTGG TTAGCTGGTT 1020  
 CATTTGGTTT TGTTGGTAT ATTTCAAATT TTGGTAACTA TTCTAAAACA TATGGCAGTA 1080  
 20 TCGCGGGTAT CATCATTTTG TTA CTATGGT TATATATCAC AAGTTTTATT ATAATTGTCG 1140  
 GnGCTGAAAT CAATGCAATC ATTCATCAGC GTAGTGTAAT TAAAGG 1186

## (2) INFORMATION FOR SEQ ID NO: 161:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7872 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

TCTTGAGCCA TCTTTTGAGC TAACTGACTA GATTGATACC CAAAAATCAT AGTTACCAAC 60  
 35 ATAAACTTTA ATTTTACCGA AGTCTAAATC AGCGATATGA GTACATACAT TATTTAAGAA 120  
 ATGACGGTCA TGCGATACTA CGATAACAGT ATTATCAAAG TTAATTAAGA AATCTTCTAA 180  
 40 CCAACTGATT GCTGGAATAT CGAGACCGTT AGTAGGCTCA TCCAGTAATA GTACGTCTGG 240  
 TTCACCGAAT AAAC TTTGCG CTAATAATAC TTTAATTTTT TGTTGTTTTT CTAATTCAGC 300  
 CATTTTTTTA TCGTGTAAG TTGGATCGAT ACCTAAACCA GATAAAAGGT TAGCAGCATC 360  
 45 AGCTTCAGCA TTCCAACCAT TCATTTCTGC AAATTCACCT TCAAGTTCAG CAGCACGGAT 420  
 ACCATCTTCA TCACTGAAAT CTGGCTTCAT ATAGATTCA TCTTTTCTT TCATAACCTC 480  
 ATAAAGACGT TCGTGACCTT TAATTACAAC ATCAAGCACG CGTTCATCTT CATAAGCATA 540  
 50 GTGGTCCTGT TTAAAACAG CTAGACGTTT ATTTTCCCT AATGAAACAT GTCCTGTTTG 600  
 AGAATCTAAT TCACCAGATA ATATTTTAA GAATGTTGAT TTACCTGCAC CATTGCGACC 660

55

	ATCTCCAAAA CGTAAACTCA CATCAGTTAC TTGTAACATG CATTTTCTCC TTTTTCAT	780
	TCGATATTCT AACGGAAGAA TTATATCATA TTATCGTCAC AGTTTCGACC TCATATAAGT	840
5	TGTAATGATA GAATGACTCA CACATGTTAT AATAATAAAG AATACAAGAA TCGAAGGAGA	900
	ATAACATGGC ATTAGACAAA GATATAGTAG GTTCTATAGA ATTCCTTGAA GTAGTAGGGT	960
	TACAAGGTTC AACTTACCTT TTTAAAGGAC CAAACGGTGA AAACGTAAAG TTAAACCAAT	1020
10	CAGAAATGAA CGATGATGAT GAATTAGAAG TAGGTGAAGA ATATAGTTTC TTCATTTATC	1080
	CAAACCGTTC AGGTGAATTA TTTGCAACTC AAAATATGCC TGATATTACG AAAGATAAAT	1140
15	ATGACTTTGC TAAAGTACTT AAAACGGATC GCGATGGGGC ACGTATAGAT GTTGGATTAC	1200
	CCCGTGAAGT GTTAGTACCA TGGGAAGATT TACCAAAAGT GAAATCACTA TGGCCACAAC	1260
	CTGGTGATTA TTTGCTAGTT ACATTACGAA TTGACCGTGA GAATCATATG TATGGACGTT	1320
20	TAGCGAGTGA ATCTGTTGTA GAAAATATGT TTACACCTGT ACACGACGAT AATTAAAAA	1380
	ACGAaGTCAT TGAAGCCAAA CCTTACCGCG TATTACGAAT TGGTAGCTTT TTATTAAGCG	1440
	AATCAGGTTA CAAAATTTTC GTACATGAAT CAGAACGTAA AGCTGAACCA AGATTAGGTG	1500
25	AATCTGTTCA AGTTAGAATT ATCGGGCATA ATGATAAAGG TGAGTTAAAT GGTTCATTTT	1560
	TACCACTTGC ACATGAACGT TTAGACGATG ACGGCCAAGT CATCTTTGAT TTAGTAGTTG	1620
	AATATGATGG TGAATTACCA TTCTGGGACA AATCAAGCCC TGAAGCGATT AAAGAAGTAT	1680
30	TCAATATGAG TAAAGGTTCA TTCAAACGTG CAATCGGTCA CTTATATAAA CAGAAGATTA	1740
	TTAATATAGA AACAGGTAAA ATCGCTTTAA CTAAAAAAGG TTGGAGTCGA ATGGACTCAA	1800
	AAGAATAATC ATTTTACAC GTGTCGTAGG ATGCGTGTTC TTTTATTCA ATATTAAATC	1860
35	GGACAGATGA AGTAGTTTTT TAAACATTCC TTTCAAAGTA AAAAATTAAA TAATTCAAAC	1920
	GAATAGGCTG GgaCATTAAAG TTCTTAGGCA ATGTAAAAAA GCTGATTTCT ATTAATTATT	1980
40	TGATGGAAAT CAGCTTTTTT GATATGTATT TTATAATGTA CAGCTCGTTG AGCTGCTATT	2040
	TTCTTATAT TAAGTGCCAT TAATACAAAA CCTAGCTCTC GTTTAACTTT ATTTATTCCT	2100
	CGAACTGACA TTCGAGTGAA aCCCAAAATA GCCTTCATAA ATCCAAAAAC AGGCTCTACA	2160
45	TAAATTTTTC TATGACTATA GATTTTTTTC GTTCTGGTT CAGAAAGCTT TTGaTTAATT	2220
	TGGGCTTTAA TGTATTTCAA AGTAAAATTA CATGTTAATA CGTAGTATTA ATGGCGAGAC	2280
	TCCTGAGGGA GCAGTGCCAG TCGAAGACAG GGGCCCCAAC ACAGAAGcTG ACATATAGTC	2340
50	AGCTTACAAC AATGTGCCGG TTGGGGTGGC TGAGACGGCA CCCTAGGAAG GGACCCGTCA	2400
	TCAAAAATTC TATTTATAGA ATTTTACAGT AATGTGACAG ACGGGCAAAG CGAagCCATT	2460

55



	CTTACTGCTG TTTTTTTAGG GATTTATGTC CCAGCCATTT TTGTATTCAT ATTTAAATTT	2580
	CGATAATTTT TCAGGAAGCA TTTTAATTTT ACTAATGAAG CAATATTTT TAGATTAACA	2640
5	AAAATTAATA TTTACATTTT CTTAACAATT TTTTATGTAA CATTTACAGT TTCTAAAAAT	2700
	GAGGTTAATA ATTCAAGGTT AAGATAAAGA TGTAATCAAT ACAAATACTA TTTGTTGTTT	2760
	ATACAGGGAG GATATTTCAA TGAAAAATG GCAATTTGTT GGTACTACAG CTTTAGGTGC	2820
10	AACACTATTA TTAGGTGCTT GTGGTGGCGG TAATGGTGGC AGTGGTAATA GTGATTTAAA	2880
	AGGGGAAGCT AAAGGTGATG GCTCATCAAC AGTAGCACCA ATTGTGGAGA AATTAAATGA	2940
15	AAAATGGGCT CAAGATCACT CGGATGCTAA AATCTCAGCA GGACAAGCTG GTACAGGTGC	3000
	TGGTTTCCAA AAATTCATTG CAGGAGATAT CGACTTCGCT GATGCTTCTA GACCAATTAA	3060
	AGATGAAGAG AAGCAAAAAT TACAAGATAA GAATATCAAA TACAAAGAAT TCAAAATTGC	3120
20	GCAAGATGGT GTAACGGTTG CTGTAAATAA AGAAAATGAT TTTGTAGATG AATTAGACAA	3180
	ACAGCAATTA AAAGCAATTT ATTCTGGAAG AGCTAAAACA TGGAAAGATG TTAATAGTAA	3240
	ATGGCCAGAT AAAAAAATAA ATGCTGTATC ACCAAACTCA AGTCATGGTA CTTATGACTT	3300
25	CTTTGAAAT GAAGTAATGA ATAAAGAAGA TATTAAAGCA GAAAAAATG CTGATACAAA	3360
	TGCTATCGTT TCTTCTGTAA CGAAAAACAA AGAGGGAATC GGATACTTTG GATATAACTT	3420
	CTACGTACAA AATAAAGATA AATTAAAAGA AGTTAAAATC AAAGATGAAA ATGGTAAAGC	3480
30	AACAGAGCCT ACGAAAAAAA CAATTcAAGA TAACTCTTAT GCATTAAGTA GACCATTATT	3540
	CATTTATGTA AATGAAAAAG CATTGAAAGA TAATAAAGTA ATGTCAGAAT TTATCAAATT	3600
	CGTCTTAGAA GATAAAGGTA AAGCAGCTGA AGAAGCTGGA TATGTAGCAG CACCAGAGAA	3660
35	AACATACAAA TCACAATTAG ATGATTTAAA AGCATTTATT GATAAAAATC AAAAATCAGA	3720
	CGACAAGAAA TCTGATGATA AAAAGTCTGA AGACAAAAAA TAATAAGACG CAATTTCAAA	3780
40	TGTGTCTTGA AACATGATTT TGATGGTGAA TCATTATTTA GAGTACAAAG CTTGATTTAT	3840
	CGAGACGCTG ATTTTGACAT TCAGTTAGTC TAcAAGCTTA TCAACTTAAA ATAGTGGTTC	3900
	ATCATTATTT TACAAATCTA ATTATTTTGG GAGTAATAGA AAGAGGTTTG ATTATGACTT	3960
45	CATCTACTAA TGTTAAAGCT TTAATCGAAA AAAATAATAA TAAAAAAGGA AAGCATAATG	4020
	ACAAAATTAT ACCAGTTATT TTAGCCGCAA TTTcAGCGAT TTCCATTTTA ACAACACTAG	4080
	GTATATTAAT CACATTGCTT TTAGAAACCA TCACTTTTTT CACCAGAATT CCAATAACTG	4140
50	AATTTCTATT TTCTACTACT TGGAATCCTA CCGGTCAGA CCCTAAGTTT GGTATCTGGG	4200
	CATTGATAAT AGGGACTTTA AAAATCACAG TTATTGCGAC TATATTTGCA GTTCCAGTCG	4260

55

## EP 0 786 519 A2

AACCGATATT AGAAATTTTA GCAGGAATAC CAACAATTGT GTTTGGTTTC TTTGCATTAA 4380  
 CCTTTGTTAC ACCAGTATTA AGATCTTTCA TACCAGGTCT TGGAGAGTTT AATGCTATAA 4440  
 5 GTCCCGGCTT AGTTGTCGGT ATTATGATTG TCCCTCTCAT CACAAGTTTG AGTGAGGATG 4500  
 CAATGGCATC TGTACCAAAT AAAATTCGAG AAGGTGCCTA TGGACTTGGA GCAACTAAAT 4560  
 TAGAAGTAGC AACTAAAGTC GTACTTCCCG CAGCAACATC AGGTATTGTA GCTTCAATCG 4620  
 10 TTCTCGCGAT TTCAAGAGCA ATTGGAGAAA CGATGATTGT ATCATTAGCG GCAGGTAGTT 4680  
 CGCCAACAGC TTCATTAAGT TTAACAAGTT CGATTCAAAC AATGACTGGA TATATTGTTG 4740  
 AGATAGCGAC AGGTGATGCA ACATTTGGAT CAAATATTTA TTACAGTATT TATGCTGTAG 4800  
 15 GGTTCACACT ATTTATCTTT ACCTTAATCA TGAATTTACT TTCTCAGTGG ATTTCTAAGC 4860  
 GTTTTAGGGA GGAGTATTAA TATGGAAACG ACAGATAATA ATAGACAATC ACTCGTCGAT 4920  
 CAACAACCTG TCCAAAAACA TTTATCATCC AGAACGGTTA AAAATAAAGT GTTCAAACCTC 4980  
 ATATTTTTAG CATGTACATT ATTAGGACTT GTCGTACTTA TTGCGTTGTT AACTCAAACA 5040  
 TTGATTAAAG GGGTAAAGTCA TTTAAATTTA CAGTTTTTCA CTAATTTTTT TTCTTCAACA 5100  
 25 CCATCTATGG CTGGCGTTAA AGGCGCGTTA ATCGGTTTAC TTTGGTTAAT GTTAAGTATC 5160  
 ATTCCATTAT CAATCATCCT AGGAATAGGT ACAGCTATAT ACTTAGAAGA ATATGCGAAA 5220  
 AACAACAAAT TTA CT CAGTT TGT TAAATC AGTATTTCCA ATTTAGCTGG TGTACCATCA 5280  
 30 GTTGTATTG GGTATTAGG TTATACTTTG TTCGTTGGTG GTGCAGGGAT TGAAGCCTTG 5340  
 AAAATGGGTA ACAGTATATT GGCAGCAGCG CTAACAATGA CCTTACTGAT ATTACCAATT 5400  
 ATTATTGTTT CAAGTCAGGA AGCAATTAGA GCTGTACCTA ACTCAGTACG CGAACTTCTT 5460  
 35 ACGGCTTAGG TGCTAATAAA TGGCAAACGA TAAGACGTGT TGTCTTACCA GCAGCGTTAC 5520  
 CTGGTATTTT AACTGGATTC ATTTGTCTC TTTCAAGAGC ACTGGGAGAA ACAGCGCCAC 5580  
 TTGTGCTAAT CGGTATACCG ACTATATTAT TGGCAACACC TAGAAGTATA TTGGATCAAT 5640  
 40 TTTCAGCATT ACCTATCCAA ATATTTACTT GGGCGAAAAT GCCTCAAGAA GAATTCCAGA 5700  
 ATGTTGCATC GGCAGGCATT ATCGTTTAC TAGTTATCTT AATCTTAATG AATGGCGTTG 5760  
 CGATTATTTT ACGTAACAAA TTTAGTAAAA AATTCTAATT TAAACAATCA ATCTCATTTA 5820  
 45 TCTATTAAAA AGGGAGTTTT AAATATGGCG CAAACACTTG CACAACTAA ACAATATCT 5880  
 CAAAGTCATA CGTTTGATGT CTCACAAAGT CATCATAAAA CACCAGATGA TACAACTCA 5940  
 50 CATTCTGTTA TATATTCAAC ACAAAATTTA GACTTATGGT ATGGCGAAAA TCATGCATTA 6000  
 CAAAATATTA ATTTAGATAT TTATGAAAAC CAAATTACTG CCATTATAGG TCCATCTGGT 6060

55

	AAAACAGCTG GTAAAATATT ATATCGAGAT CAAGACATTT TTGATCAAAA ATATTCTAAA	6180
	GAACAATTAC GTACAAATGT GGGCATGGTC TTTCAACAAC CTAATCCATT TCCAAAATCA	6240
5	ATATACGATA ATATTACTTA CGGTCCAAAG ATTACCGGTA TTAAAAATAA AAAAGTTCTT	6300
	GATGAAATCG TTGAGAAATC ATTACGTGGC GCTGCAATTT GGGATGAATT AAAGGATAGG	6360
	TtGCACACAA ATGCATATAG TTTATCCGGT GGGCAACAAC AACGTGTTTG TATCGCGCGT	6420
10	TGTTTAGCAA TTGAACCTGA AGTCATTTTA ATGGATGAAC CGACATCAGC ATTAGATCCA	6480
	ATCTCAACAT TAAGAGTAGA AGAGTTGGTT CAAGAATAA AAGAAAAGTA TACAATTATT	6540
	ATGGTtACAC ATAATATGCA ACAAGCAGCT CGTGTATCAG ATAAAACTGC ATTTTCTTA	6600
15	AATGGTTATG TCAATGAATA TGATGATACT GATAAAATTT TCTCTAACCC ATCAAACAAG	6660
	AAAACAGAAG ATTATATTTT AGGAAGGTTT GGTGATATA TAATGGCAAT AATTAGACAA	6720
20	CGATATCAGG AGCAACTTGA TGATTTAATA AAAGAATTAC GTCGGTTAGG TGCaAATGTC	6780
	TATGTGAGTA TTGaAAATGG TATAAAAtCA TTAAGTATTG aCGATAGAGG cTTTGACGA	6840
	CAAACAGTTA AAAACGATAA ACATATCAAT CAATTAAATT ATGATATTAA TGAGCGAGTT	6900
25	ATCATGTTAA TTACAAAGCA ACAGCCCAT TCGAGTGATT TGCGTATGAT GATTTCTTCA	6960
	TTAAAAATCG CCTCCGATTT AGAAAGAATA GGAGATAATG CCTCGAGTAT TGCCAATATT	7020
	CGATTGCGTA CAAAGATTAC AGATGATTAT GTGTTAACCC GTTTAAAGAC AATGGGTAAA	7080
30	TTAGCTATGT TAATGTTAAA GGACTTAGAT CAAGCATTTA AAAAGAAAGA TACCGTATTA	7140
	ATAAGAGAAA TAATTGAGCG TGATGAAGAT ATCGATGACT TATATAGTCA TATTATTAAC	7200
	GCAACGTATC TTATTGATAA CGtCCATTtG TCGCTGCACA AGCTCATTTA GCAGCAAGAC	7260
35	ATTTAGAACG TATTGGTGAT CATATTATTA ACATCGCTGA AAGTGTTTAT TTTTATTAA	7320
	CAGGTACACA TTACGAACAA TAACTTAAAG TTATTACTAT AAAATCCCTT ACGATAAATA	7380
	TATATTTCTA TTATTCATAA ACCCTCAAAA AAACCAAGAT TCTCACAATT AGTAATGTGA	7440
40	AAATCTTGGT TTATATTGTT CTACTATAAA TTGTCTCGCA TCTTAGTTAT TTGCTTGCTC	7500
	AATTTCAICT GTTAATTTTT CAACTTCATC GACTAAATCA GAAATATATT GAATTGTAGA	7560
45	TTTAAGTGGC TGTtCTGTAG TAATGTCTAC ACCTGCAATG TTTGCAAGTT CGACAGGTGA	7620
	TACACTACCA CCTTTTTTCA ATGTTTCTAA CCAAGCATCA ACAGCTGGTT GGCCTTCATT	7680
	TTTAATCTTT TGAGAAACGA CAGTTCCGAT TGTTAAGCCA GCAGAATACG TATACGAATA	7740
50	TAATCCCATA TAGTAATGAG GTTGACGCAT CCATGTTAAT TCAGCACCTT CAGTCATGTC	7800
	TACTGCATCT CCAAAAAAATT GTTTATAAAC ATTTAGCATT ATTTCATTTA ATGTnCGGCC	7860

55

## (2) INFORMATION FOR SEQ ID NO: 162:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 798 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

```

TTTTTCTTT TCTTCATTTG AAAATTGATC ATTCAGCAAT ATAAGCGTAT TTGTTAATGA      60
TTTAGGTGTT CCAATTTTCAT AATCCCACCA ATTTAAGTTG GTATTCTTGC CAGTTGTTTT      120
AGTAAAATTC TCACTTAATT CTTTTACTTT TTTATCTGGT TCTTTTCCAT ATGCATTTTT      180
ATGCAGCCAC TCAAGGGCAT CTTTCACTTT CTTCTTATTT TCGTCAGTAT TTAAAGTGGT      240
TTTAGGATTC CTCATCGCTT CTGCGATTTT CTCAATATTA CGATAGGTAC GAGTCATATG      300
AGAAGAATTA GTTTCAGGG TTTCCGCTCC TGACCACAAG TATTTCTTAC CACTTTCAGT      360
TTTCATTTCC TTGAGTAAAT TCGTCGCCTC TTTCTCTGTA GCATCAAAC TCTTCTTCAT      420
ATCTGGATTA TTCTCATCAT ACTTATCATA ACCATAGTTA ACGTCCAGCC ATGTGTTTCCT      480
CAATTTTTCa TAATCTGGCG TTTGAACATT CGTATCAGCC ACAGCGATTT GATGTTTATC      540
AACACTTCTG AATTCACCAC CATTCAAAGT AATCACACCA GCCATTAATA ACGTAATGGT      600
GGATAATTTT TGCCATTTCT TTATTCTATA TGTCATTGAc ATGTCTCCTT TTTGTGTTGC      660
GCGTGCACAA TGAATATTAT GATTAAATAA TGATTCAATT TTTCAAaATT CGTTAACGTA      720
TACAAATGAC TGTCTACTGT CAAACAATCC ACAAAGAATG TTGATGtCAT ATaACAATC      780
GATCACCCAA ATTTTCCG                                     798

```

## (2) INFORMATION FOR SEQ ID NO: 163:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5132 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

```

TACAGGTTTT ACTATAATGG ATGGTATTTT GGCTAAACGA CATTGGTTTA GTCTTCTTTT      60
TTTnACTTCC TAnATTTACA ATGGTATAAA TAATAATGCT ATATTTAGAA TGATGAGTAT      120
ACTTACTGAA ACTAAATTAA AAGTGTCTGG TTCTTTACTA AAGATAGCTG CTATCCTTGC      180

```

	AATACAAGTT CCAATGAGCG CAATTAAAAG TACTAACCCA ACGATGAAAC TCTGTTTGTC	300
	ACTTAACTCA AAGAACTAT AGATAGGATA TTTTITAATA ATCAAGCCAC CTAAATCAT	360
5	CCATAAAAAT ACGATAATTC CATAAGTCAC ATTTATAACA TACGTTATTT TTTGGTCACC	420
	AAATCGGACT AATGTATTTT GTAGAATCAG CATACCAATG ACAACACCTA AAATAACGAT	480
	ACTAGCTATA TAAAGTAAAA ATGCAATTGT CACATCAAAT GTACCCAAAT CTAAAAACCT	540
10	AGGAATTAYa AyGACTGCTA AAATAAAAGC GAAGyACAAA GTAATATAKT TATACAAACC	600
	GGTAGTAAGA CTTATCTCAG GTGATAATTG ATCAGCCATT GACTTAATCG GTGTATTAAT	660
	AATTGAACTT GTATCTTCGT TATTTTTTTC AGCCATAGTT AAATGATCTT CGAGCTCTTC	720
15	CAATAACTCT TCTACTTCTG CTTCAGTCTT ACCTCTAAAT AACAAATCAA CACGTAATTT	780
	TTCTAAAAAA TCTTGAGATT GTTTACTTAA CATCGTTTTT CCTCCAAAC AAGTTAATCA	840
20	TCCCTTTATT CAAAACCTGC CATTTTCGATT TAAATACTTT TAGTTCCTTT AAACCTGAAT	900
	CGGTAATCGT ATAGTATTTT CGCCTCGGGC CGCCATTACT AGATTTTTTT ATTGTCGTAT	960
	CAACGTATCC TTTTTTGTTC AAACGCATTA AACTGGATA AATACTACCC TCACTTATCT	1020
25	CTGGAACTC TTGATTCTTA AGTTTCGTCA TAATTTCATA TCCATACGTT TCGCCTTGGG	1080
	CAATGAGACC TAATATCGCC CCATCTAAGA GACCTTTCAT AATCTGATCT GACACTGACA	1140
	TTTTAATCAC CTACTATCTT ACATAATAAG ATAGTACATT GAGAACTTTT CGTCAACTAT	1200
30	CTTTTATTGT AAGGTAGTTG TTGTACACAT TCCTTAAATG ACTAACAAC TGTGTAATAG	1260
	GGTAATACTT ACGGAAGTAT ATTTTATTTA TGGGGGAGGA ATTAATAATG ACTACAAAAA	1320
	CAGTATTTGA TGTCATTGAT ATGGGGTTAG GATATTTAGT AAATGTGTAT GATGCTTGGA	1380
35	AAGTTGAAAA GGTACTTGAT GATTATCATA AGCCTTTTTT TAATACCATT CATTGGCAAT	1440
	TTGG <sup>5</sup> CATGT ATTAACAATT TTTGAATCGG CCTTAGCTGT TGCTGGTAAA GAGAATATTG	1500
	ATTTAAATAT CTATAGACCT TTATTCGGAA ATGGTTCGTC TCCAGATGAA TGGAAGGATG	1560
40	AAGTACCGAG TATTGAAAGG ATTTTAGAAG GTCTCCAAAC TTTACCTGAA CGTGACAGAA	1620
	ATCTAACTGA AGATGATTTA GCAATTGAAT TGAAACAGCC AATTGTCGGT TGTAAATACT	1680
45	TAGAAGAGTT ATTAGTATTA AATGCCATT CACATCCCACT TCATGCTGGT AAAATTGAAG	1740
	AGATGTCTCG TATATTAAAA AATTTAAAAT AAATATGTGC TTATTAACCG TTAACAACAC	1800
	GTTAACGGgT TTTTTATTTG TTTAAAAGGT CACTTTTTTG AATTTAATAA ACACCATCTA	1860
50	TACCAGTTCT TCACCGATTC TCGAAAAATA ATTATATTAA TGATTTCTGT AATTTAATTT	1920
	TATATTTAAT TATTACTGTA CATCTTTTGT AGTTAGCTTT ATTCTTAAAT TGAAATATGT	1980
55		

	TACTCCCTAT CGTTGTAGGT CTCCTTATTT GGGCACTTAC ACCTTTTAAA CCGGATGCTG	2100
	TGGATCCAAC AGCATGGTAT ATGTTTCGCA TATTCGTCGC GACAATCATT GCTTGTATTA	2160
5	CACAACCGAT GCCAATTGGG GCCGTCTCTA TAATTGGATT TACAATCATG GTACTCGTTG	2220
	GCATTGTTGA CATGAAAACG GCTGTCGCTG GTTTTGGTAA TAATAGCATT TGGTTAATTG	2280
	CTATGGCATT TTTCATTTCG AGAGGATTTG TGAAAACAGG TCTTGGTAGA CGTATCGCAC	2340
10	TTCATTTTCGT CAAATTATTT GGTAAAAAAA CATTAGGATT AGCATATTCT ATCGTCGGTG	2400
	TAGATTTAAT TCTAGCGCCT GCTACACCAA GTAATACCGC GCGTGCTGGT GGAATCATGT	2460
15	TCCCAATTAT CAAATCACTT TCTGAATCAT TTGGTTCGAA ACCGAAAGAC GGATCAGCAC	2520
	GCAAAATGGG TGCATTTCTT GTTTTCACAG AATTCCAAGG TAATTTAATT ACTGCGGCTA	2580
	TGTTTTTAAC TGCAATGGCC GGTAACCCCC TTGCACAAA TTTAGCATCT AGCACATCTA	2640
20	ATGTTACAT TACATGGATG AATTGGTTTC TAGCTGCTTT AGTTCCTGGA CTTGTTTCCT	2700
	TAATTGTTGT ACCTTTTATT ATTTATAAAA TTTATCCACC AACTGTTAAA GAAACACCAA	2760
	ATGCTAAGAG TTGGGCTGAA AATGAATTAG CGACTATGGG TAAATCGCT TTAGCTGAAA	2820
25	AATTTATGAT TGGTATTTTT GTCGTGCGT TAACACTATG GATTGTCGGA AGTTTCATTC	2880
	ATATTGATGC AACTTTAACG GCCTTTATTG CGCTAGcATT gTTATTATTG ACAGGCGTCT	2940
	TAACATGGCA AGACATTTTA AACGAAACAG GTGCTTGGA CACATTAGTA TGGTTCTCAG	3000
30	TATTAGTGTT AATGGCCGAC CAATTAAACA AGCTTGGATT TATTCCTTGG TTAAGTAAAT	3060
	CCATTGCTAC AAGTCTTGGT GGCTTAAGCT GGCCTATAGT CCTGGTCATT TTAATATTGT	3120
	TCTACTTCTA TTCACATTAC TTATTGCAA GTTCTACAGC ACATATCAGT GCGATGTATG	3180
35	CAGCATTACT AGgCGTTGCC ATCGCAGCCG GTGCACCACC ATTATTcAGT GCATTAATGT	3240
	TAGGTTTCTT CGGTAACCTA TTAGCTTCAA CAACACACTA TAGTAGTGGT CCAGCGCCGA	3300
40	TTCTATTCTC TTCAGGTTAC GTGACTCAA AACGTTGGTG GACAATGAAC TTAATATTAG	3360
	GTTTCGTCTA CTTTATTATC TGGATTGGTT TAGGATCACT TTGGATGAAA GTAATTGGTA	3420
	TATTTTAAAA TATTTAAATT AGCGCTCGAA TCTCATTGAT TTGGGCGCTT TTTAATTTGT	3480
45	ATTTAAAATC AACCTTTGCT AAATCAAGAC TCCCTTTTTA AAATACGTTT ATCCTTTAAA	3540
	TCATTGCGTG CTTCACTGAA AATTTGTATA AAGATTAAAG TCATTACGTA ACATCACATA	3600
	AAATACATTT CTATACTATT CCGCTTCATT GATTAAcATT ACGTATGCCC TCATAAATCA	3660
50	TCATACAAAA AACACCTTCG TTTAAATTCA TTTTAATTGC GAATTCAACG AAAGTGCCTT	3720
	ATTTCATATT TAATGTTTCA AATTTATACG TCTGTCACTG TTACTGCACA CATACCTCAG	3780

55

EP 0 786 519 A2

	TTATAGGGTT TTTGCGACCG GATGTTTCTT CAATTTAATG TATTGAGAAA GACTATATAA	3900
	CACAATACCT GTCCAAATAA ATATAAACGT AATTAAATTGA TCTATACTAA AAGGCTCTTT	3960
5	GAAAACAAAT ATGCCGAGTA CAAACATTAT TGTGGTCCA ACGTATTGAA TAAATCCTAT	4020
	TAGCGAAAGT GGAATACGTT TTGCCCCGGC TGAGAATAGG ATTAGTGGTA TTGCCGTAAT	4080
	AGCACCAGAA AATAACAACC AAAATGATGA CATGTTCAAT CCAAATGACA TCTGATGTTG	4140
10	CTGCCATAAA TAAATAACGT ATATTAGTCC AGCAGGTGCG GTAACAATAC ATTCAATCGT	4200
	AATACTGCTG ATGGCATCAA TATGTACTAC TTTTTCAT AATCCGTATG TACCAAAGGA	4260
	TAACGCTAAT ATAATAGAGA CGATTGGGAA TTCTCCAATC TTGAGCGTCA TATATAATAC	4320
15	ACCGATGAAT GCGAATAAAA TGGCTAGCCA TTCAAATTTA TTGAATCTTT CTTTAAAAA	4380
	GATAAGTGCG AGCAAAATGC TAACAAGTGG ATTTATATAA TAACCTAAAC TTGTTTGTAG	4440
20	GACGTGACCG TTCGTTACAG CCCAAATAAA TGTACCCCAA TTTAATGTAA TGACATAGCC	4500
	TGCTACGACA ATCGCTAATA GCTGAATGGG CTTGCCTAAC AATTGATTCA TATCTCGTTG	4560
	AAATGCATTG CGTTGTTTTT GTCCAACCGC GAGTATGAAA ATCATGAATA TTGCTGAAAA	4620
25	TATAATACGA AAGGCTAAAA TTTCAAATGC GCCTATTGCA TCAACGAACT GCCAATATAT	4680
	AGGTAGTATT CCCACAGAA TGTATGCACT GAGTGCTAAA AATATGCCTT TTTTATACTC	4740
	TGAATTCACC TTCAAACCTC CTTACTTTCC TAATTTTTAA TTTACTGCAT ACGCTCACTT	4800
30	GGTTATGCTA ATATAACGAT TTTACTAATA ATATTTGAT AAAGATATCA TTTGTTTAT	4860
	ATTTCCACA TTTATTCACC AACCCTAAA CAATATTAAT TTTATAAATA ATTCTGTACA	4920
	AATCAGGTA TATTGCCAGA AAGACTACCA TACAACATAA AGGATGGATA CAAATGACTT	4980
35	TACCTAAAAT TGGAAAGCCT GCAACACGCG CGCTAAATTC ACAAGGTATA TACACATTAG	5040
	AAGCAGTATC ACAATATACG AAGTCATCTC TAATGGAGAT GCATGGCGTT GGTCTAAAG	5100
40	CTATATCAAT ATTGGAACAA GCTTTATTTT AG	5132

(2) INFORMATION FOR SEQ ID NO: 164:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22243 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

AAGTAAATTA TATTATGAAT TTGCCTGTCA ATTTCTTAAA GACATTCTTA CCGGAACATA	60
---	----

	TAGAAGCAAT TAATAATGCy mAAGAAAAGA CAGCTAATAA TACCGGCTTA AAATTAATAT	180
	TTGCAATTAA TTATGGTGGC AGAGCAGAAC TTGTTTCATAG TATTAAAAAT ATGTTTGACG	240
5	AGCTTCATCA ACAAGGTTTA AATAGTGATA TCATAGATGA AACATATATA AACAAATCATT	300
	TAATGACAAA AGACTATCCT GATCCAGAGT TGTTAATTCTG TACTTCAGGA GAACAAAGAA	360
	TAAGTAATTT CTTGATTTGG CAAGTTTCGT ATAGTGAATT TATCTTTAAT CAAAAATTAT	420
10	GGCCTGACTT TGACGAAGAT GAATTAATTA AATGTATAAA AATTTATCAG TCACGTCAAA	480
	GACGCTTTGG CGGATTGACT GAGGACTAGT ATAGTATGAA AGTTAGAACC CTGACAGCTA	540
15	TTATTGCCTT AATCGTATTC TTGCCTATCT TGTAAAAGG CGGCCTTGTG TTAATGATAT	600
	TTGCTAATAT ATTAGCATTG ATTGCATTAA AAGAATTGTT GAATATGAAT ATGATTAAAT	660
	TTGTTTCAGT TCCTGGTTTA ATTAGTGCAG TTGGTCTTAT CATCATTATG TTGCCACAAC	720
20	ATGCAGGGCC ATGGGTACAA GTAATTCAAT TAAAAAGTTT AATTGCAATG AGCTTTATTG	780
	TATTAAGTTA TACTGTCTTA TCTAAAAACA GATTTAGTTT TATGGATGCT GCATTTTGCT	840
	TAATGTCTGT GGCTTATGTA GGCATTGGTT TTATGTTCTT TTATGAAACG AGATCAGAAG	900
25	GATTACATTA CATATTATAT GCCTTTTAA TTGTTTGGCT TACAGATACA GGGGCTTACT	960
	TGTTTGGTAA AATGATGGGT AAACATAAGC TTTGGCCAGT AATAAGTCCG AATAAAACAA	1020
	TCGAAGGATT CATAGGTGGC TTGTTCTGTA GTTTGATAGT ACCACTTGCA ATGTTATATT	1080
30	TTGTAGATTT CAATATGAAT GTATGGATAT TACTTGGAGT GACATTGATT TTAAGTTTAT	1140
	TTGGTCAATT AGGTGATTTA GTGGAATCAG GATTTAAGCG TCATTTCCGC GTTAAAGACT	1200
	CAGGTGCAAT ACTACCTGGA CACGGTGGTA TTTTAGACCG ATTTGACAGC TTTATGTTTG	1260
35	TGTTACCATT ATTAAATATT TTATTAATAC AATCTTAATG CTGAGAACAA ATCAATAAAC	1320
	GTAAGAGGA GTTGCTGAGA TAATTTAATG AATCTCAGAA CTCCTTTTGA AAATTATACG	1380
40	CAATATTAAC TTTGAAAATT ATACGCAATA TTAACTTTGA AAATTAGACG TTATATTTTG	1440
	TGATTTGTCA GTATCATATT ATAATGACTT ATGTTACGTA TACAGCAATC ATTTTAAAAA	1500
	TAAAGAAAT TTATAAACAA TCGAGGTGTA GCGAGTGAGC TATTTAGTTA CAATAATTGC	1560
45	ATTTATTATT GTTTTGGTG TACTAGTAAC TGTTCATGAA TATGGCCATA TGTTTTTTGC	1620
	GAAAAGAGCA GGCATTATGT GTCCAGAATT TGCGATCGGT ATGGGGCCAA AAATTTTATG	1680
	TTTTAGAAAA AATGAAACAC TTTACACTAT TAGGTTATTG CCTGTTGGTG GATATGTTTCG	1740
50	TATGGCAGGA GATGGCTTAG AAGAGCCACC AGTCGAGCCC GGTATGAACG TTAAAAATTAA	1800
	ACTTAATGAA GAAAATGAAA TAACACATAT CATATTAGAT GATCATCATA AGTTTCAACA	1860

55



	CACTGCTTAT GATAATGAAA GACATCATTT TAAAATTGCT AGAAAGTCTT TCTTTGTTGA	1980
	AAATGGTAGC TTAGTTCAAA TTGCTCCGAG AGACAGACAA TTTGCACATA AAAAGCCATG	2040
5	GCCGAAATTT TTAACATTAT TTGCGGGACC GTTATTTAAC TTTATATTAG CTTTAGTCCT	2100
	ATTATTGGT CTTGCATATT ATCaAGGcAC GCcTACGTCT ACTGTAGAAC AAGTCGCAGA	2160
	TAAGTATCCA GCTCAACAAG CAGGATTACA AAAAGGTGAT AAGATCGTCC AAATTGGCAA	2220
10	ATATAAAATA TCTGAATTTG ATGATGTTGA TAAGGCGTTA GATAAAGTTA AAGATAATAA	2280
	GACGACTGTT AAATTTGAAC GTGATGGTAA AACAAAGTCA GTTGAATTAA CACCTAAAAA	2340
15	GACTGAAAAA AAAC TGACTA AAGTAAGTTC AGAGACGAAG TATGTTCTCG GATTCCAACC	2400
	AGCGAGTGAA CATACTTTT TTAAACCAAT TGTATTCCGA TTTAAAAGCT TTTAATCGG	2460
	TAGTACTTAT ATTTTTACAG CTGTAGTAGG TATGTTGGCT AGTATATTTA CGGCGGATT	2520
20	CTCATTTGAT ATGTTAAATG GTCCGGTTGG TATTTATCAT AACGTCGACT CAGTTGTTAA	2580
	AGCGGGTATC ATTAGCTTAA TTGGTtncAC TGCGTTATTA AGTGTAAACT TAGGTATTAT	2640
	GAATTTAATT CCTATTCCTG CACTAGACGG TGGTCGTATT TTATTTGTTA TATATGAAGC	2700
25	GATTTTCAGA AAACCAGTTA ATAAAAAGC GGAAACAACG ATTATTGCTA TTGGTGCCAT	2760
	TTTCATGGTC GTTATAATGA TATTAGTAAC GTGGAATGAT ATTCGACGAT ATTTCTTATA	2820
	ATTTAGGAGG ATAAATAATT ATGAAGCAAT CCAAAGTTTT TATACCAACG ATGCGTGACG	2880
30	TGCCATCAGA AGCAGAAGCA CAAAGTCATC GTTTATTATT GAAATCGGGT TTGATAAAAC	2940
	AAAGTACAAG TGGGATTTAT AGTTATTTAC CGCTAGCAAC ACGTGTGTTA AATAATATTA	3000
	CTGCAATTGT GCGACAAGAA ATGGAACGTA TCGATTCTGT TGAAATTTTA ATGCCAGCGT	3060
35	TACAACAAGC TGAATTATGG GAAGAATCAG GACGTTGGGG TGCATATGGC CCAGAATTAA	3120
	TGCGTTTACA AGATAGaCAT GGAAGACAAT TTgCATTAGG TCCaACACAT GAAGAATTAG	3180
40	TTACATCAAT AGTAAGAAAT GAATTGAAAT CATACAAACA ATTACCGATG ACATTATTCC	3240
	aAATTCAATC TAAATTCCGT GATGAAAAGA GACCACGTTT TGGTTTayTC GTGGGCGTGA	3300
	ATTTATTATG AAAGATGCAT ATTCATTCCA TGCTGACGAG GCATCATTAG ATCAAACGTA	3360
45	TCAAGATATG TATCAAGCGT ATAGCCGTAT TTTTGAGAGA GTTGGCATTa ACGCAAGACC	3420
	AGTAGTTGCA GATTCAGGTG CTATAGGCGG TAGCCATaCA CATGAATTTA TGGCATTAAg	3480
	TGCTATCGGT GAGGATACAA TCGTTTACAG TAAAGAAAGT GATTATGCTG CTAACATCGA	3540
50	AAAAGCAGAA GTCGTTTACG ArcCAaATcA TaAGCATACT ACTGTGCAAC CTTTAGAAAA	3600
	AATTGAAACA CCAATGTTA AGACTGCGCA AGAATTGGCA GACTTCTTAG GTAGACCAGT	3660
55		

	GCGTGGCCAT CATGAAATTA ATGACATTAA ATTAAAATCT TATTTTCGGCA CAGATAATAT	3780
	TGAATTAGCA ACACAAGACG AAATTGTTAA TTTAGTTGGT GCAAATCCTG GTTCACTAGG	3840
5	TCCTGTAATT GATAAAGAAA TCAAAATTTA TGCAGATAAT TTTGTGCAAG ATTTAAATAA	3900
	TTTAGTTGTC GGTGCTAACG AAGATGGTTA TCACTTAATT AATGTAAATG TAGGTAGAGA	3960
	CTTCAACGTT GATGAATATG GCGATTTCCG TTTTATTTTA GAAGGCGAAA AGTTAAGTGA	4020
10	TGGTTCAGGC GTTGACATT TTGCTGAAGG TATTGAAGTT GGTCAAGTAT TCAAATTGGG	4080
	TACTAAGTAT TCAGAATCAA TGAATGCTAC ATTCTTAGAT AACCAAGGAA AAGCTCAATC	4140
	TTTAATTATG GGTGTTACG GAATTGGAAT TTCTAGAACG CTAAGTGCAG TTGTTGAACA	4200
15	AAATCACGAT GATAATGGAA TTGTTTGGCC TAAATCAGTT ACTCCGTTTG ATTTACATTT	4260
	AATTTCTATT AATCCTAAGA AAGATGATCA ACGAGAACTA GCAGATGCAC TATATGCTGA	4320
	ATTTAATACT AAATTTGATG TGTGTGACGA TGATCGTCAG GAACGTGCAG GTGTTAAATT	4380
20	TAATGATGCC GATTTAATTG GTTTACCACT GCGAATTGTT GTTGGTAAAC GTGCATCGGA	4440
	AGGTATTGTA GAAGTTAAAG AACGTTTAAC AGGTGATAGC GAAGAAGTTC ACATTGATGA	4500
25	CTTAATGACT GTCATTACAA ATAAATATGA TAACCTAAAA TAATTAAGAT CGAATGAATT	4560
	ATAAGAGTAG GAAAAAGCTG AAAGAAATCT GATGCTTATG TCCTGCTCTT ATTATTTTGT	4620
	ATATAATGAT TATTCGATGA AAAATGACTG AAGACATAGT ATAATTAAAG ATAAATTTGT	4680
30	TTTAACAATA TAATGATTAG CCAAATATAA AGCATTTAAT TTTCTATCAT TACTATGCTC	4740
	ACATAATCTA AATATTGTTT GAACACGTAA AAGTAATTTT TATTTAAGGT GGTAATTGTC	4800
	TTGGCAATGA CAGAGCAACA AAAATTTAAA GTGCTTGCTG ATCAAATTAA AATTTCAAAT	4860
35	CAATTAGATG CTGAAATTTT AAATTCAGGT GAACTGACAC GTATAGATGT TTCTAACAAA	4920
	AACAGAACAT GGAATTTTCA TATTACATTA CCACAATTCT TAGCTCATGA AGATTATTTA	4980
	TTATTTATAA ATGCAATAGA GCAAGAGTTT AAAGATATCG CCAACGTTAC ATGTCGTTTT	5040
40	ACGGTAACAA ATGGCACGAA TCAAGATGAA CATGCAATTA AATACTTTGG GCACTGTATT	5100
	GACCAAACAG CTTTATCTCC AAAAGTTAAA GGTCAATTGA AACAGAAAAA GCTTATTATG	5160
	TCTGGAAAAG TATTAAGAGT AATGGTATCA AATGACATTG AACGTAATCA TTTTGATAAG	5220
45	GCATGTAATG GAAGTCTTAT CAAAGCGTTT AGAAATTGTG GTTTTGATAT CGATAAAATC	5280
	ATATTCGAAA CAAATGATAA TGATCAAGAA CAAACTTAG CTTCTTTAGA AgCACaTATT	5340
50	CAAGAAGAAG ACGAACAAAG TGCACGATTG GCAACAGAGA AACTTGAAAA AATGAAAGCT	5400
	GAAAAAGCGA AACAACAAGA TAACAACGAA AGTGCTGTCTG ATAAGTGTCA AATTGGTAAG	5460

55

GCAATAGAGG GTGTCATTTT TGATATAAAC TTAAAAGAAC TTAAAAGTGG TCGCCATATC 5580  
 GTAGAAATTA AAGTGACTGA CTATACGGAC TCTTTAGTTT TAAAAATGTT TACTCGTAAA 5640  
 5 AACAAAGATG ATTTAGAACA TTTTAAAGCG CTAAGTGTTG GTAAATGGGT TAGGGCTCAA 5700  
 GGTGCTATTG AAGAAGATAC ATTTATTAGA GATTTAGTTA TGATGATGTC TGATATTGAA 5760  
 GAGATTAAAA AAGCGACAAA AAAAGATAAG GCTGAAGAAA AGCGTGTAGA ATTCCACTTG 5820  
 10 CATACTGCAA TGAGCCAAAT GGATGGTATA CCCAATATTG GTGCGTATGT TAAACAGGCA 5880  
 GCAGACTGGG GACATCCAGC CATTGCGGTT ACAGACCATA ATGTTGTGCA AGCATTTCCTA 5940  
 GATGCTCACG CAGCAGCGGA AAAACATGGC ATTAAATGA TATACGGTAT GGAAGGTATG 6000  
 15 TTAGTTGATG ATGGTGTTC CATTGCATAC AAACCACAAG ATGTCGTATT AAAAGATGCT 6060  
 ACTTATGTTG TGTTCGACGT TGAGACAACT GGTTTATCAA ATCAGTATGA TAAATCATC 6120  
 GAGCTTGACG CTGTGAAAGT TCATAACGGT GAAATCATCG ATAAGTTTGA AAGGTTTAGT 6180  
 20 AATCCGCATG AACGATTATC GGAAACGATT ATCAATTTGA CGCATATTAC TGATGATATG 6240  
 TTAGTAGATG CCCCTGAGAT TGAAGAAGTA CTTACAGAGT TTAAAGAATG GGTTGGCGAT 6300  
 25 GCGATATTCG TAGCGCATAA TGCTTCGTTT GATATGGGCT TCATCGATAC GGGATATGAA 6360  
 CGTCTTGGGT TTGGACCATC AACGAATGGT GTTATCGATA CTTTAGAATT ATCTCGTACG 6420  
 ATTAATACTG AATATGGTAA ACATGGTTTG AATTTCTTGG CTAAAAAATA TGGCGTAGAA 6480  
 30 TTAACGCAAC ATCACCGTGC CATTATGAT ACAGAAGCAA CAGCTTACAT TTTCATAAAA 6540  
 ATGGTTCAAC AAATGAAAGA ATTAGCGTA TTAATCATA ACGAAATCAA CAAAAAATC 6600  
 AGTAATGAAG ATGCATATAA ACGTGCAAGA CCTAGTCATG TCACATTAAT TGTACAAAAC 6660  
 35 CAACAAGGTC TTAAAAATCT ATTTAAAATT GTAAGTGCAT CATTGGTGAA GTATTCTAC 6720  
 CGTAACCTC GAATTCCACG TTCATTGTTA GATGAATATC GTGAGGGATT ATTGGTAGGT 6780  
 ACAGCGTGTG ATGAAGGTGA ATTATTTACG GCAGTTATGC AGAAGGACCA GAGTCAAGTT 6840  
 40 GAAAAAATTG CCAAATATTA TGATTTTATT GAAATTCAAC CACCGGCACT TTATCAAGAT 6900  
 TTAATTGATA GAGAGCTTAT TAGAGATACT GAAACATTAC ATGAAATTTA TCAACGTTTA 6960  
 ATACATGCAG GTGACACAGC GGGTATACCT GTTATTGCGA CAGGAAATGC AACTATTG 7020  
 45 TTTGAACATG ATGGTATCGC ACGTAAAATT TTAATAGCAT CACAACCCGG CAATCCACTT 7080  
 AATCGCTCAA CTTTACCGGA AGCACATTTT AGAACTACAG ATGAAATGTT AAACGAGTTT 7140  
 50 CATTTTTATG GTGAAGAAAA AGCGCATGAA ATTGTTGTGA AAAATACAAA CGAATTAGCA 7200  
 GATCGAATTG AACGTGTTGT TCCTATTAAA GATGAATTAT ACACACCGCG TATGGAAGGT 7260

55

	CTGCCTCAAA TCGTAATTGA TCGATTAGAA AAAGAATTAA AAAGTATTAT CGGTAATGGA	7380
	TTTGCGGTAA TTTACTTAAT TTCGCAACGT TTAGTTAAAA AATCATTAGA TGATGGATAC	7440
5	TTAGTTGGTT CCCGTGGTTC AGTAGGTTCT AGTTTTGTAG CGACAATGAC TGAGATTACT	7500
	GAAGTAAACC CGTTACCGCC AACTATATT TGTCCGAACGT GTAAAACGAG TGAATTTTTC	7560
	AATGATGGTT CAGTAGGATC AGGATTTGAT TTACCTGATA AGACGTGTGA AACTTGTGGA	7620
10	GCGCCACTTA TTAAAGAAGG ACAAGATATT CCGTTTGAAA CATTTTTAGG ATTTAAGGGA	7680
	GATAAAGTTC CTGATATCGA CTTAAACTTT AGTGGTGAAT ATCAACCGAA TGCCCAT AAC	7740
15	TACACAAAAG TATTATTTGG TGAGGATAAA GTATTCCGTG CAGGTACAAT TGGTACTGTT	7800
	GCTGAAAAGA CTGCTTTTGG TTATGTTAAA GGTTATTTGA ATGATCAAGG TATCCACAAA	7860
	AGAGGTGCTG AAATAGATCG ACTCGTTAAA GGATGTACAG GTGTTAAACG TACAACCTGGA	7920
20	CAGCATCCAG GGGGTATTAT TGTAGTACCT GATTACATGG ATATTTATGA TTTTACGCCG	7980
	ATACAATATC CTGCCGATGA TCAA AATTCA GCATGGATGA CGACACATTT TGATTTCCAT	8040
	TCTATTCATG ATAATGTATT AAACTTGAT ATACTTGAC ACGATGATCC AACCAATGATT	8100
25	CGTATGCTTC AAGATTTATC AGGAATTGAT CCAAAAACAA TACCTGTAGA TGATAAAGAA	8160
	GTTATGCAGA TATTTAGTAC ACCTGAAAGT TTGGGTGTTA CTGAAGATGA AATTTTATGT	8220
	AAAACAGGTA CATTTGGGGT ACCAGAATTC GGTACAGGAT TCGTGCCTCA AATGTTAGAA	8280
30	GATACAAAGC CAACAACATT TTCTGAATTA GTTCAAATCT CAGGATTATC TCATGGTACA	8340
	GATGTGTGGT TAGGCAATGC TCAAGAATTA ATTA AAACCG GTATATGTGA TTTATCAAGT	8400
	GTAATTGGTT GTCGTGATGA TATCATGGTT TATTTAATGT ATGCTGGTTT AGAACCATCA	8460
35	ATGGCTTTTA AAATAATGGA GTCAGTACGT AAAGGTAAAG GTTTAACTGA AGAAATGATT	8520
	GAAAAGATGA AAGAAAATGA AGTGCCAGAT TGGTATTTAG ATTCATGTCT TAAAATTAAG	8580
40	TACATGTTCC CTAAAGCCCA TGCAGCAGCA TACGTTTTAA TGGCAGTACG TATCGCATAT	8640
	TTCAAAGTAC ATCATCCACT TTATTACTAT GCATCTTACT TTACAATTCG TGCGTCAGAC	8700
	TTTGATTTAA TCACGATGAT TAAAGATAAA ACAAGCATTG GAAATACTGT AAAAGACATG	8760
45	TATTCTCGCT ATATGGATCT AGGTAAAAAA GAAAAAGACG TATTAACAGT CTTGGAAATT	8820
	ATGAATGAAA TGGCGCATCG AGGTATTCGA ATGCAACCGA TTAGTTTAGA AAAGAGTCAG	8880
	GCGTTCGAAT TTATCAATTGA AGGCGATACA CTTATTCCGC CGTTCATATC AGTGCCTGGG	8940
50	CTTGGCGAAA ACGTTGCGAA ACGAATTGTT GAAGCTCGTG ACGATGGCCC ATTTTATCA	9000
	AAAGAAGATT TAAACAAAAA AGCTGGATTA TCTCAGAAAA TTATTGAGTA TTTAGATGAG	9060

55

	GAAATAATCA AGGTATTTAT TTAATGCGTA TGGCGTAGTC AAAGAAATAC AAAATTGTTG	9180
	CTGGACACAA AATTATGCCC GTATTTCTTT TCAATGTCTT ACGAGTCTAT TCAAATGTAA	9240
5	TGGTGAAATA AAGGAACAAA CTTTACAAAG AATCTCTGAT TAATAGTGAA GTCATTTGTT	9300
	TCAAGCATAA ACTTATGCTA TAATTAAGTT GCTTAAAAAT TAGTGAAGTC AGGCAGAAGA	9360
	GTGGGAGATT CCCGCTCTTT TCTATTTGCC AAAAAGGGAG GCCTGTATGA GTAAAATTAC	9420
10	AGAACAAGTA GAAGTGATTG TTAAACCAAT TATGGAAGAC TTGAATTTTG AACTTGTAGA	9480
	CGTTGAATAT GTCAAAGAGG GTAGAGATCA TTTTCTTAGA ATCTCTATTG ATAAAGAAGG	9540
15	TGGCGTAGAT TTAAATGATT GTACGCTAGC TTCTGAAAAA ATAAGTGAAG CTATGGATGC	9600
	AAATGATCCT ATTCCTGAAA TGTATTATTT AGACGTAGCG TCACCTGGTG CAGAACGTCC	9660
	AATTAAAAAA GAACAAGATT TCCAAAATGC AATAACTAAA CCTGTATTTG TTTCTTTATA	9720
20	TGTACCAATT GAAGGTGAAA AGGAATGGTT AGGCATTTTA CAAGAAGTCA ATAATGAAAC	9780
	AATTGTAGTA CAAGTTAAAA TCAAAGCAAG AACGAAAGAT ATAGAGATAC CGAGAGACAA	9840
	AATAGCAAAA GCACGTCACG CAGTTATGAT TTAACGTGAT GAGGAGGAAA AAACGTGTCA	9900
25	AGTAATGAAT TATTATTAGC TACTGAGTAT TTAGAAAAAG AAAAGAAGAT TCCTAGAGCA	9960
	GTATTAATTG ATGCTATTGA AGCAGCTTTA ATTACTGCAT ACAAAAAGAA TTATGATAGT	10020
	GCAAGAAATG TCCGTGTGGA ATTAAATATG GATCAAGGTA CTTTCAAAGT TATCGCTCGT	10080
30	AAAGATGTTG TTGAAGAAGT ATTTGACGAC AGAGATGAAG TGGATTTAAG TACAGCGCTT	10140
	GTTAAAAACC CTGCATATGA AATTGGTGAT ATATACGAAG AAGATGTAAAC ACCTAAAGAT	10200
	TTTGGTCGTG TAGGTGCTCA AGCAGCGAAA CAAGCAGTAA TGCAACGTCT TCGTGATGCT	10260
35	GAACGTGAAA TTTTATTTGA AGAATTTATA GACAAAGAAG AAGACATACT TACTGGAATT	10320
	ATTGACCGTG TTGACCATCG TTATGTATAT GTGAATTTAG GTCGTATCGA AGCTGTTTTA	10380
	TCTGAAGCAG AAAGAAGTCC TAACGAAAAA TATATTCCTA ACGAACGTAT CAAAGTATAT	10440
40	GTTAACAAAG TGGAACAAAC GACAAAAGGT CCTCAAATCT ATGTTTCTCG TAGCCATCCA	10500
	GGTTTATTAA AACGTTTATT TGAACAAGAA GTTCCAGAAA TTTACGATGG TACTGTAATT	10560
45	GTTAAATCAG TAGCACGTGA AGCTGGCGAT CGCTCTAAAA TTAGTGTCTT CTCTGAAAAC	10620
	AATGATATAG ATGCTGTTGG TGCATGTGTT GGTGCTAAAG GCGCACGTGT TGAAGCTGTT	10680
	GTTGAAGAGC TAGGTGGTGA AAAAATCGAC ATCGTTCAAT GGAATGAAGA TCCAAAAGTA	10740
50	TTTGTA AAAA ATGCTTTAAG CCCTTCTCAA GTTTTAGAAG TTATTGTTGA TGAAACAAAT	10800
	CAATCTACAG TAGTTGTTGT TCCTGATTAT CAATTGTCAT TAGCGATTGG TAAAAGAGGA	10860

55

EP 0 786 519 A2

	GATGCGCGTG AAGCGGGTAT CTATCCAGTA GTTGAAGCTG AAAAAGTAAC TGAAGAAGAT	10980
	GTTGCTTTAG AAGATGCTGA CACAACAGAA TCAACCGAAG AGGTAAATGA TGTTTCAGTT	11040
5	GAAACAAATG TAGAGAAAGA ATCTGAATAA TAGGTTGGAG TGAAGTATCT ATGAAAAAGA	11100
	AAAAAATTCC GATGCGAAAA TGTATTCTTT CAAATGAAAT GCATCCCAAA AAAGATATGA	11160
	TTCGTGTTGT TGTTAATAAA GAAGGCGAAA TCTTTGCGGA TGTTACTGGA AAGAAACAAG	11220
10	GCCGTGGCGC ATATGTTTCT AAAGATGTTG CTATGTTTGA AAAAGCACAA CAAAAAGAAA	11280
	TTTTAGAAAA ATATTTTAAA GCATCTAAAG AGCAATTGGA TCCTGTTTAC AAAGAAATTA	11340
	TTAGATTAAT TTATAGAGAA GAGATCCCAA AATGAGTATA GATCAAATAT TAACTTTTTT	11400
15	AGGATTAGCA ATGAGAGCTG GTAAAGTAAA AACAGGTGAA TCAGTCATTG TTAATGAGAT	11460
	TAAAAAAGGA AATTTGAAGC TCGTTATTGT TGCAAATGAT GCGTCTGATA ATACAGCTAA	11520
20	ATTAATTACA GATAAATGTA AGAGTTACAA AGTTCCATTG AGAAAGTTTG GAAATCGAAA	11580
	TGAATTGGGA ATAGCACTTG GAAAAGGTGA GCGTGTTAAT GTAGGGATTA CTGACCCAGG	11640
	CTTTGCTAAA AAGTTGCTAT CAATGATAGA TGAATATCAT AAGGAGTGAT TATATGAGTA	11700
25	AACAAAGAAT TTACGAATAT GCGAAAGAAT TAAATCTAAA GAGTAAAGAG ATTATAGATG	11760
	AGTTAAAAAG CATGAATATT GAGGTTTCAA ATCATATGCA AGCTTTGGAA GATGACCAAA	11820
	TTAAAGCATT AGATAAAAAG TTCAAAAAG AACAAAAGAA CGACAATAAA CAAAGCACTC	11880
30	AAAATAATCA CCAAAAATCA AACATCAAA ACCAAAATAA AGGGCmACAA AAAGATAACA	11940
	AAAAGAATCm ACAACAAAAT AATAAAGGCA ACAAAGGCAA TAAAAAGAAT AATAGAAATa	12000
	ATAAGAAAAA TAACAAGAAT AATAAACCAC AAAATCAACC AGCTGCTCCA AAAGAAATAC	12060
35	CATCAAAAGT GACATATCAA GAAGGTATTA CAGTAGGCGA ATTTGCGGAT AAATTAAATG	12120
	TTGAATCATC AGAAATTATC AAAAAATTAT TCTTACTTGG TATTGTTGCT AATATCAATC	12180
	AATCATTAAA TCAAGAAACA ATCGAATTAA TTGCCGATGA TTATGGCGTT GAGGTTGAAG	12240
40	AAGAAGTTGT GATTAATGAA GAAGACTTAT CAATCTATTT CGAAGACGAA AAAGATGATC	12300
	CAGAGGCAAT TGAGAGACCA GCAGTTGTAA CAATTATGGG ACATGTTGAC CATGGTAAAA	12360
	CGACTTTATT AGATTCAATT CGTCATACAA AAGTTACAGC AGGTGAAGCA GGCGGAATCA	12420
45	CTCAACATAT TGGTGCATAT CAAATTGAAA ACGATGGCAA AAAAATCACT TTCTTAGATA	12480
	CACCGGGACA TGCTGCATTT ACAACGATGC GTGCGCGTGG TGCaCAAGTA ACAGATATTA	12540
50	CTATTTTAGT AGTAGCAGCT GACGATGGTG TTATGCCACA AACAAATTGAA GCAATTAAAC	12600
	ATGCTAAAGA AGCagAAGTA CCAATTATTG TTGCAGTAAA TAAAATTGAT AAACCAACTT	12660

55

## EP 0 786 519 A2

	GCGGCGAAAC AATTtTCGTc CACTTTCTGC ATTAAGTGGT GATGGTATCG ACGATTTATT	12780
	AGAAATGATA GGATTAGTTG CAGAAGTTCA AGAACTTAAA GCAAATCCTA AAAACCGTGC	12840
5	TGTTGGTACA GTTATCGAAG CTGAATTAGA TAAATCACGT GGTCTTCTG CATCATTATT	12900
	AGTACAAAAC GGTACATTAA ATGTTGGTGA TGCGATTGTA GTTGGTAATA CTTACGGCCG	12960
	TATCCGTGCA ATGGTTAATG ACTTAGGTCA AAGAATCAAA ACGGCTGGTC CATCAACGCC	13020
10	TGTTGAAATT ACAGGTATTA ATGATGTGCC ACAAGCTGGG GATCGCTTTG TTGTATTTAG	13080
	TGATGAAAAA CAAGCTCGTC GTATTGGTGA ATCAAGACAC GAAGCTAGCA TTATACAACA	13140
	ACGTCAAGAA AGTAAAAATG TTTCATTAGA TAACCTGTTT GAACAAATGA AACAAAGGTGA	13200
15	AATGAAAGAT TTAAACGTTA TTATTAAAGG TGATGTTCAA GGTCTCTGTG AAGCTTTAGC	13260
	TGCATCATTa ATGAAAATTG ATGTTGAAGG CGTAAATGTT CGTATCATTC ATACAGCGGT	13320
20	TGGTGCAATT AATGAGTCAG ACGTGACACT TGCTAATGCC TCAAATGGTA TTATCATTGG	13380
	TTTCAATGTT CGTCCAGACA GTGGTGCAAA ACGTGCTGCA GAAGCTGAAA ATGTTGATAT	13440
	GCGTTTACAC AGAGTTATTT ATAATGTTAT CGAAGAAATT GAATCAGCGA TGAAAGGTTT	13500
25	ACTTGATCCA GAATTTGAAG AACAAAGTTAT CGGACAAGCT GAAGTTCGTC AAACATTCAA	13560
	AGTTTCTAAA GTTGGTACTA TTGCTGGATG TTATGTTACT GAAGGTAAAA TTACGCGAAA	13620
	TGCTGGTGTA CGTATTATTC GTGATGGTAT TGTTCAATAT GAAGGCGAAT TAGATACACT	13680
30	TAAACGTTTC AAAGATGATG CTAAGGAAGT TGCAAAAGGT TATGAATGTG GTATTACAAY	13740
	TGAAAACTAC AATGACCTTA AAGAAGGCGA TGTTATCGAA GCATTTGAAA TGGTTGAAAT	13800
	TAAGCGTTAA TTAAATAAAT TACAAGCTAA AAGTATAGTT AAGATTGATA TGCTCCCTAT	13860
35	AAATATTGCA CTTTTTAAGT GTCTACTTTA TAGGGAGCAT ATTTGATACT AGCTTTTGGT	13920
	TTTPTATTAG AATAGATTAC CTATTAAAAG TTACGTTATA TGGACATGAT TTTGTATAAA	13980
	ATTTTGTGGT GGCCTAGAAT GATTTTAAAT GACAAAATAT AATGTCGACT ATTATTGGAA	14040
40	AATTTTCTGT TGaAATGCCT ATCTTACGGC AAACCTTTATT TGATTTTATA GGCTTAATTT	14100
	ATTAAAATAA CGTGTGAGCT AAAATAATTG TTTAAGCATT GTTACACTAA AAAATGCAAA	14160
45	TAACAATTGA ACTTAAAGAT AAAGAGGTGA CAAGAATGAG CAGTATGAGA GCAGAGCGTG	14220
	TTGGTGAACA AATGAAGAAG GAATTAATGG ATATCATCAA CAATAAAGTC AAAGATCCTC	14280
	GAGTTGGTTT TATTACAATT ACAGATGTTG TTTTAACAAA TGATTTATCG CAGGCTAAAG	14340
50	TATTTTAAAC TGTATTAGGT AACGATAAAG AAGTAGAAAA TACATTTAAA GCACTTGATA	14400
	AAGCAAAGG CTTCATTAAG TCTGAATTAG GTTCTAGAAT GCGATTACGT ATTATGCCGG	14460

55

EP 0 786 519 A2

	AAGATTTACA CAAACAAGAT AGATAATTTA GTGTTAGGTA TCTGGAAAAT GTTTGATAAT	14580
	TTCTTAATAT CGGTATATTA ACATTAAACA GTTAATACAT AGATGTGTAG AAATAGTTAA	14640
5	CATTTTCCAG TTTTPTTATG AATAAATTTA GTTGATACGC TATTAAAATA TATTTTAAAA	14700
	AAGAAGGTGA CTATATGTAT AATGGGATAT TACCAGTATA TAAAGAGCGC GGTTTAACAA	14760
	GTCATGACGT TGTATTCAAA TTGCGTAAAA TATTA AAAAC TAAAAAATA GGTCAACACGG	14820
10	GTACGCTTGA TCCCGAAGTT GCAGGCGTGT TACCGGTATG TATAGGTAAT GCAACGAGAG	14880
	TTAGTGATTA TGTATGGAT ATGGGCAAAG CTTATGAAGC AACTGTATCG ATAGGAAGAA	14940
	GTACAACGAC TGAAGATCAA ACGGGTGATA CATTGGAAAC AAAAGGTGTA CACTCAGCAG	15000
15	ATTTTAATAA GGACGATATT GACCGATTGT TAGAAAGTTT TAAAGGTATC ATTGAACAAA	15060
	TTCCGCCGAT GTACTCATCC GTCAAAGTAA ATGGTAAAA ATTATATGAA TATGCGCGTA	15120
20	ATAATGAAAC AGTTGAAAGA CCAAAGCGTA AAGTAAATAT TAAAGACATT GGGCGTATAT	15180
	CTGAATTAGA TTTTAAAGAA AATGAGTGTC ATTTTAAAT ACGCGTCATC TGTGGTAAAG	15240
	GTACATATAT TAGAACGCTA GCAACTGATA TTGGTGTGAA ATTAGGCTTT CCGGCACATA	15300
25	TGTCGAAATT AACACGAATC GAGTCTGGTG GATTTGTGTT GAAAGATAGC CTTACATTAG	15360
	AACAAATAAA AGAACTTCAT GAGCAGGATT CATTGCAAAA TAAATTGTTT CCTTTAGAAT	15420
	ATGGATTAAA GGGTTTGCCA AGCATTAAAA TTAAAGATTC GCACATAAAA AAACGTATTT	15480
30	TAAATGGGCA GAAATTTAAT AAAATGAAT TTGATAACAA AATTAAAGAC CAAATTGTAT	15540
	TTATTGATGA TGATTCAGAA AAAGTATTAG CAATTTATAT GGTACACCTT ACAAAGAAT	15600
	CAGAAATTAA ACCTAAAAAA GTCTTTAATT AAAGGAGATA GAATTTATGA AAGTCATAGA	15660
35	AGTGACACAT CCTATACAAT CTAAACAGTA TATTACAGAG GATGTTGCAA TGGCATTCCG	15720
	ATTTTTCGAT GGCATGCATA AAGGTCATGA CAAAGTCTTT GATATATTAA ACGAAATAGC	15780
	TGAGGCACGC AGTTTAAAAA AAGCGGTGAT GACATTTGAT CCGCATCCGT CTGTCGTGTT	15840
40	GAATCCTAAA AGAAAACGAA CAACGTATTT AACGCCACTT TCAGATAAAA TCGAAAAAAT	15900
	TAGCCAACAT GATATTGATT ATTGTATAGT GGTTAATTTT TCATCTAGGT TTGCTAATGT	15960
	GAGCGTAGAA GATTTTGTG AAAATTATAT AATTAAAAAT AATGTAAAAG AAGTCATTGC	16020
45	TGGTTTTGAT TTTACTTTTG GTAAATTTGG AAAAGGTAAT ATGACTGTAC TTCAAGAATA	16080
	TGATGCGTTT AATACGACAA TTGTGAGTAA ACAAGAAAT GAAAATGAAA AAATTTCTAC	16140
50	AACTTCTATT CGTCAAGATT TAATCAATGG TGAGTTGCAA AAAGCGAATG ATGCTTTAGG	16200
	CTATATATAT TCTATTAAAG KCACTGTAGT GCAAGGTGAA AAAAGGGGAA GAACTATTGG	16260

55



	TGCTGTTAGT ATTGAAATCG GCACTGAAAA TAAATTATAT CGAGGGGTAG CTAACATAGG	16380
	TGTAAAGCCA ACATTTTCATG ATCCTAACAA AGCAGAAGTT GTCATCGAAG TGAATATCTT	16440
5	TGACTTTGAG GATAATATTT ATGGTGAACG AGTGACCGTG AATTGGCATC ATTTCTTACG	16500
	TCCTGAGATT AAATTTGATG GTATCGACCC ATTAGTTAAA CAAATGAACG ATGATAAATC	16560
	GCGTGCTAAA TATTTATTAG CAGTTGATTT TGGTGATGAA GTAGCTTATA ATATCTAGAG	16620
10	TTGCGTATAG tTATATAAAC AATCTATACC ACACCTTTTT CTTAGTAGGT CGAATCTCCA	16680
	ACGCCTAACT CGGATTAAGG AGTATTCAAA CATTTTAAGG AGGAAATTGA TTATGGCAAT	16740
	TTACAAGAA CGTAAAAACG AAATCATTA AGAATACCGT GTACACGAAA CTGATACTGG	16800
15	TTCAACAGAA GTACAAATCG CTGTACTTAC TGCAGAAATC AACGCaGTAA ACGAACACTT	16860
	ACGTACACAC AAAAAAGACC ACCATTACG TCGTGGATTA TTAAAAATGG TAGGTCGTCG	16920
20	TAGcATTTaT TAAACTACTT ACGTaGTAAA GATATTCAAC GTTACCGTGA ATTAATTAAA	16980
	TCACTTGGTA TCCGTCGTTA ATCTTAATAT AACGCTTTTG AGGTTGGGGC ATATTTATGT	17040
	TCCAACCTTA ATTTATATTA AAAAAGCTTT TTACAAATAT TAACATTAT TATATGTTAA	17100
25	GCTAATATTG AGTGAATAAT AAGGTTACAA TGAGATAAAG ATGATATAAG TACACCTAGA	17160
	GTAATAATCA AGATATTAAA AATAAAGTAT GTTTTTTTAA AAAATATAAC TTATATTTAT	17220
	ACTGATAAGG GTGGGACGAT AAGTCTATTT TGTAAATAAT AGATGGATAT CCCGCTCTCT	17280
30	TTTTTTCCAA TTCAATATTT TATACTAAT ATTAATAATC GATAATAAAT GATATGATAT	17340
	AACTATTAGA TTCAAGAGAG GAGATTTATA ATGTCTCAAG AAAAGAAAGT TTTTAAACT	17400
	GAATGGGCAG GAAGATCTTT AACGATTGAA ACAGGGCAAT TAGCTAAACA AGCAAATGGC	17460
35	GCTGTATTGG TTCGTTATGG AGATACAGTC GTGTTATCGA CGGCAACTGC ATCAAAAGAA	17520
	CCTCGTGATG GAGATTTCTT CCCATTAACA GTGAACATG AAGAAAAAAT GTACGCTGCG	17580
	GGTAAAATTC CTGGTGGATT TAAAAAGAGA GAAGGACGTC CTGGTGACGA TGCAACATTA	17640
40	ACTGCGCGAT TAATTGATAG ACCAATTAGA CCTTTATTCC CTAAAGGATA TAAGCATGAT	17700
	GTTCAAATTA TGAACATGGT ATTAAGTGCA GATCCTGATT GTTCACCACA AATGGCTGCA	17760
45	ATGATTGGTT CATCTATGGC GCTTAGTGTG TCGGATATTC CATTCCAAGG GCCAATCGCC	17820
	GGTGTAAATG TGGGTTATAT TGACGGTAAA TATATCATT ACCCAACAGT AGAAGAAAAA	17880
	GAAGTTTCTC GTTTAGACCT TGAAGTAGCT GGTCAATAAG ATGCGGTAAA CATGGTAGAG	17940
50	GCAGGCGCTA GTGAGATTAC TGAACAAGAA ATGTTAGAGG CGATTTTCTT TGGTCATGAA	18000
	GAGATTCAAC GTTTAGTTGA TTTCCAACAA CAAATCGTCG ACCACATTCA ACCTGTAAAA	18060

55

	GAAGAAAAAG	GACTTAAAGA	AACAGTTTTA	ACATTTGATA	AACAACAACG	AGATGaAAAT	18180
	CTTGATAACT	TAAAAGAAGA	AATCGTCAAT	GAATTTATCG	ATGAAGAAGA	TCCAGAGAAT	18240
5	GAaTTACTTA	TTAAAGAAGT	TTATGCAATT	TTAAATGAAT	TAGTGAAAGA	AGAAGTTCGA	18300
	CGTTTAATTG	CAGATGAAAA	AATTAGACCA	GACGGCCGTA	AACCTGATGA	AATCCGTCCA	18360
	TTAGATTCTG	AAGTTGGTAT	TTTACCTAGA	ACGCATGGTT	CAGGTCTATT	TACACGTGGT	18420
10	CAGACTCAAG	CACTTTTCAGT	TTTAACATTA	GGTGCTTTAG	GCGATTATCA	ATTAATTGAT	18480
	GGTTTAGGAC	CTGAAGAAGA	AAAAAGATTG	ATGCATCATT	ACAACTTCCC	GAATTTTTCA	18540
15	GTAGGTGAAA	CTGGTCCAGT	ACGTGCGCCA	GGTCGTCGTG	AAATTGGACA	TGGTGCGTTA	18600
	GGTGAAAGAG	CATTAAAATA	TATTATTCCT	GATACTGCTG	ATTTCCCAT	TACAATTCGT	18660
	ATTGTAAGTG	AGGTACTTGA	ATCAAATGGT	TCATCATCTC	AAGCGTCAAT	TTGTGGATCA	18720
20	ACATTAGCAT	TAATGGATGC	GGGCGTACCG	ATTAAAGCAC	CAGTTGCTGG	TATTGCTATG	18780
	GGCCTTGTTA	CACGTGAAGA	TAGCTATACG	ATTTTAACTG	ATATCCAAGG	TATGGAAGAT	18840
	GCATTAGGTG	ATATGGACTT	TAAAGTCGCT	GGTACTAAAG	AAGGTATTAC	AGCAATCCAA	18900
25	ATGGATATTA	AAATTGACGG	TTTAACGCGT	GAAATTATCG	AAGAGGCTCT	AGAACAAGCG	18960
	AGACGTGGTC	GTTTAGAAAT	AATGAATCAT	ATGTTACAAA	CAATTGATCA	ACCACGTACT	19020
30	GAATTAAGTG	cTTACGCGCC	AAAAGTTGTA	ACTATGACAA	TTAAACCAGA	TAAGATTAGA	19080
	GATGTTATCG	GACCTGGTGG	TAAAAAAATT	AACGAAATTA	TTGATGAAAC	AGGTGTTAAA	19140
	TTAGATATTG	AACAAGATGG	TACTATCTTT	ATTGGTGCTG	TTGATCAAGC	TATGATAAAT	19200
35	CGTGCTCGTG	AAATCATTGA	GGAAATTACA	CGTGAAGCGG	AAGTAGGTCA	AACTTATCAA	19260
	GCCACTGTGA	AACGTATTGA	AAAATACGGT	GCGTTTGTAG	GCCTATTCCC	AGGTAAAGAT	19320
	GCGTTGCTTC	ACATTTTACA	AATTTCAAAA	AATAGAATTG	AAAAAGTGGA	AGATGTATTA	19380
40	AAAATCGGTG	ACACAATTGA	AGTTAAGATT	ACTGAAATTG	ATAAACAAGG	TCGAGTAAAT	19440
	GCTTCACATA	GAGCATTAGA	AGAATAATAT	TTAAAGTCAT	ATGACGACAA	TGTATCGTCA	19500
45	TGTGATTTTT	TTATGCCACT	TTTTACGAAG	TGACCCGTTT	TGAATTTGTT	GTATTGAACA	19560
	TTTTAAAACG	CTTTATTATT	TTGTGTGCAA	CTGTTAATTA	TCCTGTATGT	ATAGTGATT	19620
	ATAGTGATCA	TCAAGTGTTC	TTTAACTTAT	AATGAATAGT	GAGTTTATAT	ATGGACGGGT	19680
50	AACAAATTTA	GGAGGTAAGA	TTTTGAGTTT	AATAAAGAAA	AAGAATAAAG	ATATTCGCAT	19740
	TATACCATTA	GGCGGTGTTG	GCGAAATTGC	TAAAAATATG	TATATCGTTG	AAGTAGACGA	19800
	TGAAATGTTT	ATGTTAGATG	CTGGACTTAT	GTTTCCAGAA	GACGAAATGC	TAGGTATTGA	19860
55							

EP 0 786 519 A2

	CCTTACACAC	GGACATGAGC	ACGCGATTGG	TGCAGTGAGT	TATGTTTTAG	AACAATTAGA	19980
	TGCACCAGTA	TATGGATCTA	AATTGACAAT	AGCGTTAATT	AAAGAAAATA	TGAAAGCCCCG	20040
5	TAATATTGAT	AAAAAAGTTC	GCTACTATAC	AGTTAATAAT	GATTCAATTA	TGAGATTCAA	20100
	AAACGTGAAT	ATTAGTTTCT	TTAATACGAC	ACACAGTATT	CCTGATAGTT	TAGGTGTTTG	20160
	TATTCACACT	TCATATGGTG	CCATTGTGTA	TACAGGTGAA	TTTAAGTTTG	ACCAAAGTTT	20220
10	ACATGGACAT	TATGCACCAG	ATATTAAACG	TATGGCAGAG	ATTGGTGAAG	AAGGCGTATT	20280
	TGTCTTAATC	AGTGATTCTA	CTGAGGCAGA	GAAACCTGGA	TATAATACTC	CGGAAAATGT	20340
	GATTGAACAT	CATATGTATG	ATGCTTTTGC	AAAAGTGCAG	GGTCGCTTGA	TAGTTTCATG	20400
15	TTATGCTTCG	AACTTTATAC	GTATTCAGCA	AGTTTTAAAT	ATTGCTAGCA	AGCTAAATCG	20460
	TAAAGTGTC	TTTTTAGGAA	GATCACTTGA	AAGTTCATTT	AATATTGCTC	GTAAAATGGG	20520
20	GTATTTTCGAC	ATTCTCTAAAG	ATTTGCTAAT	TCCTATAACA	GAAGTTGATA	ATTATCTCTAA	20580
	AAATGAAGTG	ATAATTATAG	CTACTGGTAT	GCAAGGAGAA	CCTGTAGAAG	CCTTAAGTCA	20640
	AATGGCGCAA	CATAAGCATA	AAATTATGAA	TATCGAAGAA	GGCGATTCTG	TATTTTTAGC	20700
25	AATTACGGCT	TCTGCTAATA	TGGAAGTTAT	CATTGCGAAT	AcATTAAATG	AGCtTgTtAC	20760
	GnCTGGCGCA	CATATTATTC	CAAATAACAA	AAAGATTCAT	GCTTCAAGTC	ATGGTTGCAT	20820
	GGAAGAATTA	AAAATGATGA	TTAATATTAT	GAAACCTGAA	TACTTTATTC	CTGTACAAGG	20880
30	TGAATTTAAA	ATGCAGATAG	CACATGCGAA	GCTAGCAGCT	GAAGCAGGTG	TTGCACCAGA	20940
	AAAGATTTTC	CTTGTGGAAA	AAGGAGATGT	CATTAATTAC	AACGGTAAAG	ATATGATATT	21000
35	AAATGAAAAG	GTAAATTCAG	GAAATATTTT	AATAGATGGC	ATTGGTATTG	GGGATGTAGG	21060
	AAATATCGTG	TTGAGAGACC	GTCATCTTTT	AGCAGAAGAT	GGTATCTTTA	TTGCTGTTGT	21120
	AACGTTAGAT	CCTAAAAATA	GACGTATAGC	TGCGGGACCT	GAAATTCAAT	CTCGTGGGTT	21180
40	TGTATATGTA	CGTGAAAGTG	AAGACTTATT	ACGTGAAGCA	GAAGAGAAAG	TACGTGAAAT	21240
	AGTAGAGGCT	GGTTTACAAG	AAAAACGCAT	AGAATGGTCT	GAAATTAAAC	AAAATATGCG	21300
	TGATCAAATT	AGTAACTAT	TATTCGAAAG	TACAAAACGT	CGTCCTATGA	TTATTCCAGT	21360
45	AATTTCTGAA	ATTTAATCAA	AAAGTCATTA	ACATAAAAGA	GGTCAGAACA	AGTCACTGAA	21420
	ATATAATGGT	TGTCATGGAC	AATTTACTTA	TATTTTATGA	TAGTCAATTG	AAGGGGTAAC	21480
50	GATTAATCTG	TTATCTTAAG	TAAATTGATA	CATAGATGAT	ATTGTTCTAA	CCTCTTTTCAT	21540
	CGTCTGTTTG	GACTACATAT	TCTAAACATC	AAATAGGAAA	TTATATATAA	TAACGTCGTT	21600
55	TTAACTAAGG	CAACATAAGG	AGGTGCGTCA	ATTGGCACAA	GCAAAAAAGA	AATCGACAGC	21660

GATACGTTAT GTCATAGCTA TTTTAGTAGT TGTATTAAATG GTGTTGGGTG TTTTCCAATT 21780  
 AGGAATAATA GGTGCTCTAA TTGACAGCTT CTTTAATTAT TTATTGGGT ACAGTAGATA 21840  
 5 TTTAACATAT ATTTTAGTAC TCTTAGCAAC TGGTTTTATT ACATACTCTA AACGTATTCC 21900  
 TAmAaCTAGA CGAACGGCTG GTTCGATTGT ATTGCAAATT GCATTGCTAT TTGTATCACA 21960  
 GTTAGTTTTT CATTTTAATA GTGGTATCAA AGCTGAAAGA GAACCTGTAC TTTCTTATGT 22020  
 10 GTATCAGTCA TACCAACACA GTCATTTCCC AAATTTTGGT GCGGGTGTAT TAGGCTTTTA 22080  
 TTTATTAGAG TTAAGCGTAC CTTTAATTTT ATTATTTGGT GTATGTATTA TTAATTTTTT 22140  
 ATTATTATGC TCAAGTGTTA TTTTATTAAC AAACCATCAA CATCGTGAAG TGCAGAAAGT 22200  
 15 TGCACTGGAA AATATAAAAG CTTGGTTTGG TTCATTTAAT GAA 22243

## (2) INFORMATION FOR SEQ ID NO: 165:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5510 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

TTATTAATna TTAATATTTT TATTTTAAA AATAAAGCGA GGAGCTATCA ATGGAACAAA 60  
 30 TTAATTCTGC ACAAATAAT AGAATTAAAC AAGCGAACAA GCTAAAaAG AACGTGAGA 120  
 GGGATAAAC TGGATTAGCT TTAATTGAAG GTGTGCATTT AATTGAAGAA GCTTATCAAA 180  
 35 GTGGAATTGT AATTACACAA TTATTTGCAA TTGAACCGGC AAGATTAGAT CAGCAAATTA 240  
 WCGCATACGC GCAAGAAGTT TTTgAAATAA ACATGAAAGT TGCTGAATCT TTATCAGGTA 300  
 CAGTGACACC ACAAGGGTTT TTCGCAATCA TTGAGAAGCC GCATTATGAT ATTTCTAAAG 360  
 40 CACAACAAGT ATTGCTCATC GATCGTGTTT AAGATCCTGG AAATTTAGGC ACATTAATTA 420  
 GAACTGCGGA TGCTGCTGGA ATGGATGCTG TAATAATGGA GAAGGGTACG ACAGATCCTT 480  
 ATCAAGATAA AGTGTTCGA GCGAGTCAAG GTAGTGTTTT CCATTGCCA GTTATGACAC 540  
 45 AAGATCTCGA TACGTTTATT ACTCAATTTA ATGGTCCTGT TTATGGTACA GCACTTGAAA 600  
 ACGCAGTGgC ATACAAAGAA GTTACTTCAA GTGATTCCTT TGCATTACTA TTAGGTAATG 660  
 50 AGGGAGAAGG TGTTAATCCT GAATTATTAG CACATACTAC ACAAATTTA ATCATACCTA 720  
 TTTATGGTAA AGCTGAAAGT TAAATGTAG CGATTGCAGG TAGTATTTTA CTTATCATT 780  
 TGAAAGGTTG ACCGTGTTGA AAGTTTCCG ATATAATTAT AATTAATTGT TTAACAGAAC 840

55

	ATAAATAATT	GTTTTAGGGA	GAATAATCGT	GACTGCAAGT	TATTCCAATT	ATTTAAAGTC	960
	TTTTCACCTT	TTTGGTTACT	TAAAGAGATT	TAAGTCGGAA	AGACAATCCG	TTATCAATAT	1020
5	TAAACAAGTG	TATGCTTAGG	CATAAATTTG	GGTGGTACCA	CGGAAATGAC	TTTCGTCCCT	1080
	TATTTTTTAA	GAGGATGAAA	GTCTTTTTTT	AGTTAAACAA	CAAATATGAT	AAATAGAAAA	1140
	TGAATAGTTC	GAATAGGGAG	GTCAGTGACA	TATGTCTGAA	CAACAAACAA	TGTCAGAGTT	1200
10	AAAACAACAA	GCGCTTGTAG	ATATTAATGA	AGCAAATGAT	GAACGTGCAC	TGCAAGAAGT	1260
	TAAAGTGAAA	TACTTAGGTA	AAAAAGGGTC	AGTTAGCGGA	CTAATGAAAT	TGATGAAGGA	1320
15	TTTGCCGAAT	GAAGATAAAC	CTGCGTTTGG	TCAAAAAGTG	AATGAATTGC	GTCAAACAAT	1380
	TCAAAATGAA	TTAGATGAAA	GACAACAGAT	GTTAGTTAAA	GAAAAATTAA	ATAAGCcaAT	1440
	TGGcTGAAGA	AACAATTGAT	GTATCATTAC	CAGGTCGTCA	TATTGAAATC	GGTTCAAAGC	1500
20	ATCCATTAAc	ACGTACAATA	GAAGAAATTG	AAGACTTATT	CTTAGGTTTA	GGTTATGAAA	1560
	TTGTGAATGG	ATATGAAGTT	GAACAAGATC	ATTATAACTT	CGAAATGCTG	AATTTACCTA	1620
	AATCACACCC	TGCACGTGAT	ATGCAAGATA	GTTTCTATAT	TACGGATGAA	ATTTTATTAC	1680
25	GTACGCATAC	ATCACCAGTG	CAGGCACGTa	CGATGGAATC	ACGTCATGGT	CAAGGTCCAG	1740
	TTAAAATTAT	TTGCCCTGGT	AAAGTGTATC	GTCGTGACTC	TGATGATGCG	ACACATAGTC	1800
	ATCAATTTAC	ACAAATCGAA	GGATTAGTTG	TTGATAAAAA	CGTTAAAATG	AGTGATTTGA	1860
30	AAGGTACTTT	AGAATTGTTA	GCTAAGAAAT	TATTTGGTGC	TGATCGTGAA	ATTCGTTTAC	1920
	GTCCAAGTTA	CTTCCCATTC	ACTGAACCTT	CTGTAGAAGT	TGATGTGTCA	TGTTTTAAAT	1980
35	GTAAAGGAAA	AGGTTGTAAT	GTGTGTAAAC	ACACAGGATG	GATTGAAATT	TTAGGTGCTG	2040
	GAATGGTACA	TCCTAATGTA	TTAGAAATGG	CTGGTTTTGA	TTCTTCAGAG	TACTCTGGAT	2100
	TTGCATTTGG	TATGGGACCA	GACCGTATTG	CAATGTTGAA	ATATGGTATA	GAAGATATTC	2160
40	GTCATTTCTA	TACTAATGAT	GTGAGATTTT	TAGATCAATT	TAAAGCGGTA	GAAGATAGAG	2220
	GTGACATGTA	ATGTTGATAT	CAAATGAATG	GTTGAAAGAA	TATGTAACAA	TCGATGATTC	2280
	TGTAAGTAAT	TTGGCAGAAC	GTATTACGCG	CACAGGTATT	GAAGTGGATG	ATTTAATTGA	2340
45	CTACACAAAA	GATATCAAAA	ATTTAGTTGT	CGGCTTCGTT	AAGTCAAAAG	AGAAACATCC	2400
	TGATGCTGAT	AAATTAAATG	TTTGCCAAGT	TGATATCGGA	GAAGACGAAC	CTGTACAAAT	2460
50	CGTTTGTGGT	GCACCGAACG	TTGaTGCAGG	ACAATATGTC	ATTGTTGCTA	AAGTAGGTGG	2520
	CAGATTGCCT	GGTGGTATTA	AAATTAAGCG	TGCCAAATTA	CGCGGTGAAC	GTTCAGAAGG	2580
55	TATGATTTGT	TCGTTACAAG	AAATTGGTAT	TTCAAGTAAC	TATATACCGA	AAAGTTTTGA	2640

	ATATTTAGAT GATCAAGTAA TGGAATTTGA TTTAACGCCG AATCGTGCAG ATGCTTTAAG	2760
	TATGATAGGT ACTGCTTATG AAGTTGCAGC ATTATATAAT ACAAAAATGA CTAAGCCAGA	2820
5	GACAACATCA AATGAGCTTG ATTTATCTGC AAATGATGAA CTGACTGTGA CAATTGAAAA	2880
	TGAAGATAAA GTACCATATT ATAGTGCACG TGTGTTCAC GACGTGACAA TTGAACCCTC	2940
	GCCAATTTGG ATGCAAGCAC GCTTAATAAA AGCGGGTATA CGTCCTATTA ATAATGTTGT	3000
10	TGACATTTCA AATTATGTGT TATTAGAATA CGGTCAACCA TTGCACATGT TTGATCAAGA	3060
	TGCGATTGGT TCACAACAAA TTGTTGTTTCG TCAAGCTAAT GAAGGCGAAA AAATGACAAC	3120
15	ATTAGATGAT ACAGAACGTG AATTATTAAC GAGCGATATT GTCATTACTA ATGGACAAAC	3180
	TCCAATTGCA TTAGCTGGTG TTATGGGTGG CGATTTTTCA GAAGTTAAAG AACAAACATC	3240
	AAATATAGTG ATTGAAGGTG CTATTTTTGA TCCAGTTTCA ATTCGTCATA CATCAAGACG	3300
20	TTTAAATTTA CGCAGTGAAT CATCTAGTCG TTTTGAAAAA GGAATAGCTA CTGAATTTGT	3360
	AGATGAAGCA GTCGACCGTG CATGTTATTT ATTACAAACT TATGCAAACG GAAAAGTGCT	3420
	AAAAGATAGA GTGTCTTCAG GAGAACTTGG TGCATTTATT ACACCAATCG ACATCACTGC	3480
25	TGATAAAATT AATCGCACTA TTGGATTTGA TTTGTCACAA AATGATATTG TTACTATTTT	3540
	TAATCAACTA GGGTTTGATA CAGAAATAAA TGATGATGTT ATTACAGTGC TAGTACCATC	3600
	ACGTCGTAAA GATATTACAA TTAAAGAAGA TTTAATTGAA GAAGTTGCAC GTATATATGG	3660
30	ATACGACGAT ATTCCATCAA CGTTACCTGT CTTCGATAAA GTTACTAGTG GTCAGCTAAC	3720
	TGATCGCCAA TATAAACTA GAATGGTTAA AGAAGTGTTA GAAGGTGCTG GATTAGACCa	3780
35	AGCTATTACG TATTCGTTAG TTTCTAAAGA AGATGCTACT GCaTTTTCGA TGCAACAGCG	3840
	TCAAACAATT GATTTATTGA TGCCAATGAG TGAAGCGCAT GCGTCATTAC GTCAAAGTTT	3900
	ATTACCACAT TTAATCGAAG CGGCATCATA TAATGTGGCA CGCAAAAATA AAGATGTAAA	3960
40	ATTATTTGAA ATCGGCAATG TCTTCTTTGC TAATGGAGAA GGTGAACTAC CAGATCAAGT	4020
	TGAATATTTA AGTGGTATTT TAACTGGAGA TTATGTAGTC AATCAATGGC AAGGTAAGAA	4080
	AGAAACGGTT GATTTCTATT TAGCAAAAGG TGTCGTGGAT CGAGTATCTG AAAAGTTAAA	4140
45	TCTTGAATTT AGTTATCGCC GTGCTGATAT TGaTGGATTA CATCCAGGTC GTACTGCTGA	4200
	AATCTTATTA GAGAATAAAG TTGTTGGTTT TATTGGTGAA TTACATCCAA TATTAGCAGC	4260
50	TGATAATGAT TTAAAACGTA CGTATGTTTT TGAGTTGAAT TTTGATGCAT TAATGGCTGT	4320
	GTCGGTAGGT TACATTAATT ACCAGCCAAT TCCGAGATTC CCAGGCATGT CTCGTGACAT	4380
	TGCATTAGAA GTAGATCAAA ATATTCCAGC AGCTGATTTA TTATCAACGA TTCATGCACA	4440
55		

AAAAGGTAAA AAATCAATTG CAATACGTTT AAATTATTTA GACACAGAAG AACATTGAC 4560  
 AGATGAGCGC GTTTCAAAAG TACAAGCGGA AATTGAAGCA GCATTAATTG AACAAAGGTGC 4620  
 5 TGTTATTAGA TAATGATTTA AACCCCATGT ATAAGGATAT CTGAAGTAGA TTGATATCCC 4680  
 TAACATGGGG TTTTATTTTT GGGTTCACCA ATTTGGTTCC AATGCATTTA AAAAGTCAAA 4740  
 GAGGAACAGC GGAATACAGA TGATGcTTcG CACAACtGCA TAAAAGCCTC TAATGATTAA 4800  
 10 AAATCAAAGA GGCTTTAAAA TTTTTTGGGC TTTTTCACGA TTTTtAAAAT GCTTTTTTGA 4860  
 AATGGTATCT AAACGTGAAA GACCGTATTT TTTTATAATT TTGGCGGCGA TTACATCGAC 4920  
 TTTAGCACCG GCACCTTTAG GAATCGTCAT ATTAATATTT TTTGATATTT GATCCATATA 4980  
 15 TGTAACAAAT GCGTATCGAG AAATTATGCT TGCCACTGCA ATGGCTAATG ACTTCGATTc 5040  
 TCCTTTTGTT TCAAATTTTG TTTTCTTTGG AAGTGGTATA TCTGATAATG CGTAATGGCT 5100  
 20 ATACACTTCG CGTTTTGCGA ACTGATCAAT GACGATATAG TCTAATTGAG ACGAATCAAT 5160  
 TTTTTCAGT ACATTTTGA TGGCTTCATT ATGAAGGGCA GCTTTCATTT TTACTTGAGT 5220  
 CCAGCCTTTT GCTTGCTGAA TATTATATTT TTCATTGTGT AGTGTTAATA ATGAATGTGG 5280  
 25 TATGAAAGTA ACCAATTGCT CAGCAAGTTC TACAATTTTG GTATCGGTTA ATTTTTTTGA 5340  
 ATCATCTACA CCCAAAGTTT TTAATAAGG GACATGCTCT TTGGTAACGA AAGCAGCACA 5400  
 CACAGTCAAC GGACCAAAGT AATCGCCACT TCCAGCCTCA TCACTACCAA TACAGTTAAA 5460  
 30 TTGrTCATAC ATTAAAGTTg TcCagAAAAG AATTAGCCAT ATTTnCCTTT 5510

(2) INFORMATION FOR SEQ ID NO: 166:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9623 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

GnTTTACTT ATAAATTTTA CGGGGGTAAT ATAATACTtA TTTACCTGTA ATATATGATA 60  
 45 ATTCTTCAGC GGCAGCTGCG TTGATAGTTC TATGAGAAAT GATACCTAAT CCTTTAACAT 120  
 TGGATTCTGA AATAACGATA GAACCATCAC TGTTAACTTT TTCAACAAAT GCTACATGAC 180  
 CGTAATGTTG ATCTGCACCA AATTGTCCAG CCTCAAATAC AACAGCAGCA TGACGTTTTG 240  
 50 GTGTATGACT TACTTGATAA TCACGGTATT GAGCTCGATT ATTCCAATTA TGTGCATCAC 300  
 CTAAATCACC TGAGATAGAT GTACCAAATT GTTTCATACG GTTATATACG TACCAAGTAC 360

55

	ATGAATCATC ATAATCCTTG ATAGAACGTT CATATTTATC TAAATCTGGC ATGCGTTCAT	480
	CGTCAAACG AGTTAATTGA TAGTGTTTAA TAATACTGTT TAATTTCTTA GCATAGTTTG	540
5	GATCTGTAGC ATATGTTTTA GATAAGTGTG ATGTTGCATC TTTATAAGAA TCGGCTTCCG	600
	ATTTCCATGT TGGTTTATAA ATTGTTTCGAT TGCCATCAAT ACCATTTTTA ATAAGGTCAG	660
	AGTAATCTTT TAGTGATTCT TTCGTGCTTG GATATTTTCG GAATCCAGCA TTAATACTAT	720
10	ACAATTGATT ACCATCAGCT TCTAATGTGT TAAAAGGAAC AGAATTCCTT TCaAAAGCAC	780
	CTTTGATACC GAATAAATTA TGGTTTGGTG ACWTAGCTAA AGCACTACGA CCTGAGTCAG	840
15	ATTCTAAGAT TGCTTGGGCA ATCATGACAG ACGCATAAAT ATCGTTATCT TGACCAATGC	900
	GATGTGCATC TTTAGCAATT GATTTGACAA ATTGACGTGT ATCTTTTGAG TCAACAACGT	960
	TAAATTGTCC GCTATCATCA TTGTTAGATA TACTAGGATC TGTTTCGAAT AATGATGTTG	1020
20	CACGTGTATC CTTTGTATTA ACATCGTTAT TGAATGATTG AGCAGGTTTA GATTTATGTT	1080
	TCAATTCATC TTGTGTTGGT AACTGTGGAT TCTTTGTATT AGATTTTCA TTTTGTCTT	1140
	TTTTAGATTG AGATGCATAA TCTTTTGTG TTTTCTTGC ATCTTCACTG TATTGATCCA	1200
25	AAATAGAGTC TAAAGCCGAA TCTGACATTG ATTGATTATC TTTCGATGAA GATTTTGTGAT	1260
	TTGCTTTATC GTCACCTTGCT GGTTGACTAT TTGATTGATT AGGTTGTGTT GGCTTTGGCG	1320
	AATTTGGTTG CTTATTAGAT GTACTTGGTT TTGTATTGTT TGATTTAGGT GCTTTTGTGAT	1380
30	TGTCTGCTTT ATCTGTTTA GATGATTGCG TATCAGTGTC ATTTTGTATG CTATTGTCAC	1440
	TGTTTTTATT CGAATCATTT GTTGACTTTT CGCCATTACG AGGTTGTTCG TAATCAGAAA	1500
35	TATCCGAATT TAAATTGAAT AAGTTTTGGA TTAAAGTTGT TAATGAGTAA TTATCATCGT	1560
	ATTTATTTTT GGTTAGCAAT TGGTTTATAT TGGTTTGTGG TAAATCTTA TAAATAAAAT	1620
	CAATGATATT GTTAGAGTCT GAAGTGCTGT CGTCTATAGT TTTAAATTTT TTGTCGTTAT	1680
40	TGTCTTGGTT ACTTGTATTA TTTTGTCTG CTTTATCAAT ATCTTTACTT GTAGTATCCT	1740
	TAGAAGTTTC ATCGTCATTA GATTTTTTTG AATCATGAGA TGTTGTCTTA GCTGTAGTAT	1800
	CTTTTTGAGG TGTATCAGCA TAAGCGTAG GTGAAaCTAA AGTAGGTAAT ACGAGCGTAG	1860
45	TTGATAGCAA ATAAATTAAA ATTTTATTTT TAGGCATATT TCGTATTCTC CCTTGAAAAA	1920
	TATAATAATT AAGTGTGATA ATAACTATG ATTTGTTATA ATTTATCGTA TGCTGAAAAT	1980
50	AGTTGATAGG TATCAATCGA CTAAATATCT TCCAGTAAAT TGATTATACT AATTCACAAC	2040
	GCAAAAATAA ATTAATTTAC AAAAAATATA TAAAAAATAT GAATAATTCC TACATAGGAG	2100
	TGTGACAATG AAGAACGCAT TTAAATTATT TAAATGGAT CTGAAGAAAG TAGCTAAGAC	2160
55		



	TAAC TTATGG	GCAATGTGGG	ATCCATATGG	CAACACGGGA	CACATCAAGG	TCGCAGTCGT	2280
	TAATGAAGAT	AAAGGCGACA	CAATCAGAGG	GAAAAAAGTT	AATGTCGGTA	ATACGATGGT	2340
5	TAATACACTC	AAGAAAAATA	AAAGTTTGA	TTGGCAGTTT	GTAAGTAGAG	AGAAAGCTGA	2400
	TCATGAGATA	AAAATGGGTA	AATATTTTGC	AGGTATTTAC	ATCCCATCTA	AGTTTACACA	2460
	TGAAATTACA	GGGACACTAC	GTAAGCAGCC	TCAAAAAGCA	GATGTAGAAT	TTAAGGTGAA	2520
10	TCAGAAGATT	AACGCTGTTG	CGTCTAAGCT	AACAGATACT	GGTTCGTCAG	TTGTCGTTGA	2580
	AAAAGCGAAT	GAACAATTTA	ATAAAACAGT	AACTCGAGCA	TTATTAGAAG	AAGCTAACAA	2640
15	AGCAGGTTTA	ACTATTGAAG	AAAATGTGCC	GACAAATTAAC	AAGATAAAAA	ATGCGGTATA	2700
	TTCAGCAGAT	AAAGCTTTAC	CTAAGATTAA	TGACTTTGCG	AATAAAATTG	TATATTTGAA	2760
	TAACCACCAA	GCGGATTTAG	ATAAATATGC	CAATGATTTT	AGAAAAC TAG	GAAATTATAA	2820
20	AGGTGATATT	TTAGATGCTC	AGAAAAAATT	AAACGAAGTC	AATGGTGCTA	TTCCGCAACT	2880
	TAATGAAAAG	GCTAAGTTGA	TATTAGCTTT	AAATAATTAT	ATGCCGAAAA	TTGAAAAAGC	2940
	GTTAAATTTT	GCAGCTGATG	ACGTGCCAGC	GCAGTTCCCT	AAAATTAATC	AAGGACTTAA	3000
25	CATTGCGAGT	CAAGGTATTG	ATCAAGCTAA	TGGACAGTTA	AATGATGCCA	AAGGCTTCGT	3060
	CACACAAGTT	AGAAGTAGAG	TCGGTGATTA	TCAAGATGCA	ATTTCGACGCG	CGCAAGATTT	3120
	AAATCGAAGA	AACCAGCAAC	AGATTCTCTA	AAATAGCGCG	GCGAACAACG	AAACATCAAA	3180
30	TAGTGCACCT	GCAGCTGGTA	ATGGTGTAGC	ATCAACGCCA	CCAAGTGCAC	CAAGTGGCGA	3240
	TACTGCACCA	AATAATAATG	TTACGCAAAA	TACCGCACCA	AATAGTAATA	ATGCGCCTGT	3300
35	ATCGACTACA	CCACAAAGTA	CAAGCGGGAA	AAAAGATGGT	CAAAGTTTGT	TAGATATAAC	3360
	AACAACACAA	GTCAGCACAG	CTAACGAGAA	CACACAAAAC	ATTACAGATA	AAGATGTTAA	3420
	ATCAATGGAA	GCGGCATTAA	CGGGCTCTTT	ATTATCATTA	TCAAATAATT	TAGATACCCA	3480
40	AGCGAAAGCC	GCACAAAAAG	ATAGTCAGGC	ATTACGTAAT	ATTTCGTATG	GGATTTTAGC	3540
	ATCGGACAAG	CCTTCTGATT	TTAGAGAGTC	TTTAGATAAT	GTAAAGTCCG	GTTTAGAATA	3600
	CACAACGCAA	TATAATCAAC	AATTTATCGA	TACATTAAAA	GAGATTGAGA	AGAATGAAAA	3660
45	TGTTGATTTA	TCAAAAGAAA	TTGATAAGGT	AAAAGCAGCT	AATAATCGAA	TTAATGAATC	3720
	ATTAAGGTTA	GTAAATCAAT	TAAGCAATGC	ATTAAAGAAT	GGTAGTTCAG	GAAGTGTCTG	3780
50	AGCTACTAAA	TTACTAGATC	AACTTTCAAA	ACTAGATTCA	TCATTATCAT	CATTTAGAGA	3840
	TTATGTTAAA	AAAGATCTTA	ACAGCTCTTT	AGTATCAATA	TCACAACGTA	TTATGGATGA	3900
	ATTGAACAAA	GGGCAAAC TG	CATTATCCAA	TGTTCACTCT	AAATTAAATA	CAATTGATCA	3960
55							

	AACAGTATTA CCAAGTATTG AACACAATA CATTAGTGCT GTTAAAAATG CTCAAGCAAA	4080
	CTTCTCGAAA GTGAAAAGTG ATGTAGcTAA AGCTGCTAAC TTTGTGCGCA ATGACTTACC	4140
5	ACAGTTAGAA CAGCGATTAA CTAATGCGAC AGCAAGTGTG AATAAAAAAT TACCAACGTT	4200
	ATTAAATGGT TATGATCAAG CGGTAGGATT ACTAAATAAA AATCAGCCAC AAGCGAAAAA	4260
10	GGCTTTATCA GATTTAGCTG ATTTTCTCA AAATAAATTG CCTGATGTTG AAAAAGATTT	4320
	GAAAAAAGCG AATAAAATTT TCAAGAAATT AGACAAAGAT GATGCAGTCG ACAAATTAAT	4380
	CGACACACTT AAGAATGATT TGAAAAAGCA AGCGGGTATT ATTGCAAATC CTATTAATAA	4440
15	GAAGACTGTT GATGTTTTCC CAGTTAAGGA TTATGGTTCA GGTATGACAC CATTCTATAC	4500
	TGCACTGTCA GTATGGGTAG GTGCACTCTT GATGGTAAGT TTATTAACGG TTGATAATAA	4560
	ACATAAGAGT CTAGAGTCAG TCTTAACGAC AAGACAAGTG TTCTTAGGTA AGGCAGGATT	4620
20	CTTTATAATG CTTGGTATGT TGCAAGCACT CATTGTATCG GTTGAGATT TGTTAATCCT	4680
	AAAAGCAGGA GTTGAGTCAC CTGTATTATT TGTACTTATA ACGATTTTCT GTTCGATTAT	4740
	TTTCAACTCA ATCGTATATA CGTGCATATC ATTACTTGGT AACCCAGGTA AAGCCATTGC	4800
25	AATCGTATTG CTTGTATTAC AAATTGCAGG TGGTGGGGGA ACATTCCCAA TTCAAACCTAC	4860
	GCCACAATTT TTCCAAAACA TTTCGCCATA CTTACCATT TCGTATGCAA TTGATTCATT	4920
30	ACGTGAAACA GTAGGCGGTA TTGTTCGGGA AATCCTAATT ACAAATTAAT TTATATTAAC	4980
	GTTATTTGGT ATAGGATTCT TCGTTGTAGG TTTAATTTTA AAACCTGTAA CAGATCCATT	5040
	GATGAAGCGC GTATCTGAAA AAGTTGACCA AAGTAACGTT ACAGAATAAA AATTAAATCC	5100
35	ACACATTAGG GTTATAGCTC CTTAATGTGT GGATTTTTAT GTTTTTAGAC AGAAGAGATA	5160
	GTAATTTCTG TCTTTTATGG GACGGTTGTT ATCATTGCTA TTATCCAGGA TGACTTACTA	5220
	TAGC <sup>1</sup> ACTAAT ATTACCGACA AAGTGAATAT CCTCGTCTTC CGTAGTTAAA ATAAAGCTAG	5280
40	AAC <sup>1</sup> TTTTTG GATGTCATAG TGCTTATCGT TTA <sup>1</sup> CTGTAA AGTACCAGTA CCATCGATAA	5340
	TTGTAACTAA GCAATAAGCA TGTGGTTTAT TGAATTTTAA ATCTCCATGA ATATCCCAT	5400
	TATATACTGC AAAATATTGA TTATCTACAA ATTGAGTTAC AGTGTGTGTG TCGATGTGAG	5460
45	TTGTTATAGG AGTAGTATTT GGTTCATGAT TGCCTAATTC AATCACATCT TTACTTTGCT	5520
	CTAAGTGCAA ATCACGCAAT TGACCATTTT GATCTCGTCT ATCATAGTCA TAAATACGGT	5580
50	ATGTCGTATC GGAGGATTGT TGTGTCTCTA AAATTAAAAT ACCCGAACCA ATGGCATGGA	5640
	CAGTGCCAGC AGGAACATAA TAAAAGTCAC CGGGCTTAAC AGGTATACGT TTGAAAAGAC	5700
	TGTCAAATTC ATGATTATCA ATCATGTCTA TTAACGTCTG TTTATTATGT GCATGTACGC	5760
55		

	GTTCGCCTTC GTGTTTTAAA GCGTAGTCAT CATCTGGGTG AACTTGAACA GATAATTTAT	5880
	CATTGGCATC TAATACTTTA GTTAGCAGAG GGAAACTATC TCGTGAATCA TTATCGAATA	5940
5	ATTCACGATG TTGTGACCAA AGTTGATCTA GGGTCATATC CTTGTATGGA CCATTGATAA	6000
	TTGTATTAGG ACCATTTGGA TGTGCAGAAA TTGCCAGCA TTCACCAGTT GTTTCATTAG	6060
	GGATATCATA GTTAAATGCT TTTAATGCAT GACCGCCCCA AATTCTGTCT TTAAAAACGG	6120
10	GTTGTAAAAA TAATGCCATA GTTAAAACTC CTCTATATTT TCATTAATAA GTTATAAATT	6180
	TCTGTAGTAC TGTTTGCATT AATTAGTGAT TGGCGTGTCT CATCATTCAT TAACGCTTTA	6240
15	GATAAGCGCT GAAGTATTTT TAAATGTGTA TCCTGACTGT TGTTTGGTAC GGCAATTAAG	6300
	AATATCAATT GAGGTAGACT ACCATCTAGA CTGTCCCATT TAACACCATG ATTATTTTTTC	6360
	ATAACAGCTA CAATCGGTTG TTTTACAACA TCAGACTTTG CATGTGGAAT GGCCACGTTT	6420
20	ATGCCAATAG CTGTCGTAGm tCcATTTCAC GTTCTAGTAT TGCATTTTTT AAATGCGATG	6480
	TGTGCTCTAC ATAACGGCAA ATTTTAAGTT TATGAATCAA CATATCAATT GCTTCGTTTT	6540
	GAGACATGTC GTGATCAGTA ATTATCATAG TTTGTGTATC AAAAACATGA GAAGGTTTAT	6600
25	TGAGATGTGA ATGTTTCGCG GTGTTATCTA CATTGTCAAC CTCTGTATCA TGTGTGTAA	6660
	TATCTGTATC ATGAAGTTGC GTGTGTTGCG CTGGTGCATC TACTGCTATA ACTGGTGTAT	6720
	TGCGTTTTAA TAATAGTACA GTAGTCATTG TGACAAGACT ACCTACTATC ACTGCAAAGA	6780
30	TAAACCATAA TACATGATCA ATACCACCTA ATACAGCCAC GATTGGACCT CCATGTGCGA	6840
	CTCTATCGCC GACACCACCA ATGGCTGCAA TGACTGATGC AATCATTGCA CCAATGATGT	6900
35	TTGCAGGTAT AATGCGCAAT GGATCTTGGG CTGCGAAAGG AATAGCACCT TCAGTAATAC	6960
	CAAATAGTCC CATAGTGAAG GAAGCCTTAC CCATTTCTCT TTCGGAATGA TTGAATTTAT	7020
	ACTTTTGAAC AAACGTTGCT AAACCTAAAC CGATTGGTGG TGTACATACA GCAACTGCGA	7080
40	CCATACCCAT AACGGCGTAA TTACCTTCAG CAATAAGTGC TGAGCCAAAT AAAAATGCTA	7140
	CCTTGTTTAC TGGACCGCCC ATATCGAAGG CAATCATCGC ACCTATAATC ATCGCAAGTA	7200
	TAATAATATT AGCACCTTGC ATACTTTTTA ACCAGGTTGT TAATGCCTCA AAAATATTAG	7260
45	AAATTGGTGC ACCGATTAAA AATATAAATA TCAATCCTAC AACGACCGAT GAAATAATGG	7320
	GAATAATAAT GATAGGCATA ATTGGTGCCA TTGCTTTTGG AACTTTAATA TCTTTAATCC	7380
	ACTTTGCGAT ATAACCTGCT AAGAAACCAG CAACAATACC ACCTAAAAAT CCTGCGCCTG	7440
50	CATCACTGCC ATAAAACTA CCGTCAGCAG CGATAGCGCC GCCAATCATA CCAGGAACAA	7500
	GACCGGGcTT GTCAGCGATA CTAACAGCGA TATATCCAGC TAGTATTGGA ACCATAAATT	7560
55		

	ATCCTTTTGA TGTCGTTTCA CCGCCTAGAG TCAGCGCGAT GCGGATAAGG AGTCCACCAA	7680
	CTACGATAAA AGGAACCATA AACGATACAC CGTTCATTAA ATGTTGATAC ACCATTTGAA	7740
5	TACCATTTTT AGACTTACCG CGATCTTTTCG AATGATAATT TGTTTCAGAT TGATAAATAG	7800
	GCGCATCTTG ATTAATGATA CGTTGAATTA GACCTCTCGG ATTATGAATC CCTTCGCGAA	7860
10	CATTTTCATT AATCAACCGT TTACCAACAA ATCGGGACAG ATCAACTTGT TTATCAGCTG	7920
	CAATTATGAC ACCGTCAGCT TCTTCGATGT CTTGCGTAGT TAAAACATTT TCAGCACCAA	7980
	CACCGCCCTG TGTCTCTACT TTAATATCCA CACCCATTTT TTTTGCTACC TGCTCAAGCT	8040
15	TTTCTTGAGC CATATATGTA TGTGCAATGC CATTTGGGCA TGAGGTAATA GCTACAATTT	8100
	TCATAAAATC ATCTCCTTTT CTATATTGTA AGCGTATTCT CGATACTAAA AAAAAGAATA	8160
	ATTACCGTTA CTAGTGGCAA TTATTCTTGT AAGTATTCAA ATAAGTGTG CTTTAACTA	8220
20	TGATCATCTA AACTACATAA ATGGTTCACCT GAATCATCAT CCAAGTTAGC AATTAATTGC	8280
	ATCATTTGTT TTGTAAAAGC TTTGTCTTTA TCGGAAATCG CTAAGAAAAA GACAAGTTTG	8340
	ACATCGTGTT GTCGCCAAGG AAAAACATCT TTTGTGCGAA AAATAAGCAC ATGTGATTGT	8400
25	AAAACCTTTT CAGGATCTCC ATGAGGAATC GCCATAAAAT TACCTATGTA TGTAGAAGAT	8460
	GATTTCCTAC GCTCTAAAGC TGATTGATA TATCCTTCTA CAATCGCATG ATGTGCTTGT	8520
	AATATTTTTT GAGCTTCTTC AAAAATTTGC ACAGTATGCC GTGATTTTGT TTCAGTATTT	8580
30	ACGACAAGGA AATTGACAGT GTCCATATGA TGATGTGCTT GAACCGGATT TTGCTTTTGC	8640
	TTCAACAGT GTCTGATTTT GTGACGATCA TCTTCAGAAA ATAATGGTGC AACCTTGATA	8700
35	GTCGTCAGGT GCTTAGGAAG TATGTTTAGC GTTTGTTTAG GAATATCATG GGTGCTTATT	8760
	AATAAATCTA CATTGTCAAA GTGATAGTGT GTTATATTTT CTAGTTTAAT CGTATTTATC	8820
	ACTGACAACCT CTTCGGATAA GTTATTTATT TTAGTTTCTA AAAAATTCGA CACACCTAGA	8880
40	CCATAATAAC AAGCAATGAC TACATTTAAT TGTGTTTTGG TACGACGCTC GATGGCAGCT	8940
	TGAAAATGAA TTGTTAAAAA TGCAATTTCA TCTTCGCTCA TCTCTATATC AGTATCAATT	9000
	GCTAATTTAT CAATCGCTTC AAAAAGTGTG TTAAACACAA AGGGATAGAG TTTTPTAATC	9060
45	TCTATAACTA AAGGATTGTT TAAATAAATG TTTTGAGTGA TACGTAAATA TGCTTTACTA	9120
	AAATGATTAT ATAAATTTTG TTGTAAAATC GAATCTTCAT TGAAAGGTAC ATGAATACGT	9180
	TGCTGCATCA ATTCGATTAA GCGATCAATA TAACTTTGTA TAAATATACG TTCTATGCCA	9240
50	ATATCGAGTT TATTAAAATG ATAAGCAATA AAGAATGAAA ACATATTGAT TACTTTTTCG	9300
	TTCAAGTCAT AACCTAATCT TTCGTTGATT TGCTTAATGC AAGATTGAGA TATCAATTTT	9360
55		

AGATGAATTA AAAGCTGTTG TATTGAATA TCAGTTGTTT CAATACTATG TTGTTGAAGT 9480  
 GTCTCTTGTA TAATATGCGA AATCATCCTT TGGTGTGAAT CAGGTAATTC aTTTAAAATT 9540  
 5 AGGTCTTCAA CATGTACATG CCCTGATGAT AATTGATTTA AATGGATGAT GGCATTAGTG 9600  
 ATATCATTAT CTGTTCCATC GAC 9623

(2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1021 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

20 ACCGTGGAAC CACGCTAGT CAATCAGAAA GCGATAAAAA TGTGACTAAA TCATCTCAAG 60  
 AGGAAATCA AGCAAAAGAA GAATTACAAA GCGTTTTAAA CAAAATTAAC AAACAATCAA 120  
 GTAAGAATAA TTAAAAAATT TTGATATTGT CTATGTTTAT AGTTCACAAG CCATTCAACG 180  
 25 TATTGTAAAC TAAGGATAGT GTATTTTTTT AATAGTAATT TGTCAGGAGG TGCCTATCTA 240  
 TGGAAGAACA TTAATACGTA AGTATTGATA TTGGATCATC AAGCGTAAAA ACAATAGTAG 300  
 GCGAGAAATT TCACAAATGGT ATAAATGTGA TAGGTACAGG ACAAACCTAC ACGAGCGGTA 360  
 30 TAAAAAATGG TTTAATTGAT GATTTTGATA TTGCGCGACA AGCAATCAAA GACACAATTA 420  
 AAAAGGCATC AATCGCTTCG GGTGTTGATA TTAAAGAAGT TTTCTGAAA TTACCTATCA 480  
 35 TTGAACGGA AGTTTATGAT GAATCAAAATG AAATCGACTT TTATGAGGAT ACAGAAATCA 540  
 ACGGTTTACA TATCGAAAAA GTATTAGAAG GTATTAGAGA AAAAAATGAT GTGCAAGAAA 600  
 CAGAAAGTAAT TAATGTGTTC CCGATTCTGT TTATAGTCGA TAAAGAAAAT GAGGTTTCAG 660  
 40 ACCCTAAAGA ATTAATTGCC AGACATTCAT TAAAGGTTGA AGCAGGCGTA ATTGCTATTC 720  
 AAAAATCGAT TTTAATTAAT ATGATTAAAT GCGTAGAAGC ATGTGGTGTT GATGTATTAG 780  
 ATGTTTACTC TGATGCATAT AACTATGGTT CAATCCTAAC AGCTACTGAA AAAGAGTTAG 840  
 45 GTGCATGTGT CATTGATATT GGTGAAGACG TTACGCAAGT TGCTTTTTAT GAACGCGGTG 900  
 AATTAGTAGA TGCTGATTCT ATCGAAATGG CAGGCGGTGA TATTACaGAC GATaTTGCAC 960  
 aAGGrTTaAA CACTTCTnAT GAACTGCTG nAAAAAGTTA AACACCAATn TGGTCATGCA 1020  
 50 T 1021

(2) INFORMATION FOR SEQ ID NO: 168:

(A) LENGTH: 7963 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

10	TAATCTATTA TAAAACTGT CCATACCCTT TGATTACCTT CTCTTCAGGT ACAGGCCACA	60
	CTTGAGGCCA TAAGCCATAT GCTTGCTGTG AATAAAATTG TGCCATTGTG AACAAATATAA	120
	TATATACAAA TAAACACCCA ATAATTGCTG TCACTAATGG ATATGATAAC CAAACCATTAA	180
15	ATAAACTGC AATAATTACT AACCTAAAGA TAATATTAAA TCGTCTCTC CCTCTTATAA	240
	AGCTTCTAAT AAATAAGAAT AAATACATCG CATTAGAGTT AAATTTACTA CCCTTTGGAA	300
	CTGGTAAAAG TATATCTAGA TAACTTCTTC TGA CTGCAGA TTCTTTCAA TGT TTTACAT	360
20	CGGTGAACAT ATTAACAAAT TTATAATAAT TCATATGATG TCGATGTTG ATTGCAATCA	420
	TTTTCTCCCA AGGATACAAA AAGCCTGGTT TATATTTTTT AACTAAAAAT TCTATTAACA	480
	CAGGCAAAGC AACCATCACA AATGCGATGT ACCATTTTGG AGCTAATAGT AAGTAATATG	540
25	TTAGAGCAAA GGTGATGAAT GATATTAAAT TAACTTGCCA TGT TTTAAGT CCCGATTGAT	600
	ACCATTGCCA TCTTAAGCGT AAACCAACAT ATGGAAAAAT TAATGCACTG ACTCCAAAAC	660
30	AAATATAAAA TGCCACATTA TGTTGATTAA TATGTAAAA CAACGGGAAC ATTACAATAA	720
	CAATAATGAG TTGGATTAAT ATGCGCGCAA AGTAACTATA TAAAATCGCA TGACGCATAA	780
	ATTGAGACAT GTGTTTTTCA AATGGTAATA AAAAGATTTT ATCCgCTTCT TTTAACAGTG	840
35	GTCsCmTTGG AAAAATAGrT GTCAACGCAA CAATCACTGC TGCTATTaAT GAAAAATTGa	900
	TATTCGTTGG AATATGTTTT AACCATTcAC CATATCCaR AATAAATGCA CCCAGCAAAA	960
	TAAGTAAAAA GACCATGAAA TGACCATTAA ATATAAACTT ATTATAATAA TTTTtCTCTT	1020
40	TACGAAGGGC ATGTAATCTT TTATTAAATA ATGTGGTgGC TTGGTTACGC ATGTACATCT	1080
	CCACCTTGCG TCACATGAAT ATATATATCG TCTAATGTTT GATTATGTAA GCCAGTTTGT	1140
	TGTCTCAATG CTTCTAAATC TCCAAATGCA ACGACTTCAC CTTCTCTAG TATGaTAAAA	1200
45	CGATCACAGT AACGTTCAGC TGTGCTAAA ATATGTGTAC TCATTAGAAC GGTCTACCT	1260
	TCGTTTTTCT TTTCAACCAT TAAATCTAAC ATGGATTGAA TTCCTAATGG ATCTAGGCCA	1320
50	AGGAATGGTT CGTCTATAAT ATACAATTCG GGATTAAACGA TAAACGCACA AATAATCATG	1380
	ACTTTTTGTT TCATCCCCTT AGAAAAATGA CTCGGAAAAA CTTTCAACTC ATTTTCTAAA	1440
	CGGAATGTCT TTAATAATGG CATTGCTCGA TTCATCGTTT CATCACGATC AATATCATAT	1500

55

	TCCGGAATAT AAGATAACTT TCTTCTATAA GCCTCTATGT CATCATTAAT GTTGATATCT	1620
	GAAATTGATA GAGATCCTTC CATAGGTGTA AGCAATCCTA GCATATGTTT AATCGTTGTA	1680
5	CTCTTACCAG CGCCATTAAG GCCAATAAGT CCAACAATTT CGCCTTTGTT TAATTCAAAA	1740
	TTTATATCTT TAATTACAGG GCGTTTTCCA TATCCACCTG TAAGCTGTTC TACTTTAACT	1800
	GTCATAAGGC ACCTCCATGA CTTATATTGT ACCAAAAATT ATAAAATGCT CATATTAAAT	1860
10	ACACATGTCC TAATATCGAA TTTTTCGCGA CAATGTTATA ATGAATGGTA ATACTAGTTG	1920
	AAAAGGAGTG TAGTCATCAT GTCAGAAACA ATTTTCGGCA AAATTTTAAC TGGAGAAAT	1980
15	CCTAGCTTTA AAGTATATGA AGACGATTAT GTCTATGCCT TTTTAGATAT ATCACAAGTT	2040
	ACTAAAGGAC ATACGTTATT AATTCCTAAA AAAGCTTCTG CTAATATCTT TGAACTGAT	2100
	GAAGAAACAA TGAAACATAT CGGTGCAGCA TTACCTAAAG TAGCAAATGC TATTAAGCGT	2160
20	GCATTTAATC CTGATGGTTT AAACATTATT CAAAATAATG GTGAGTTTGC AGATCAATCT	2220
	GTATTTTATA TTCAATTTCCA CTTAATTCCT CGATACGAAA ATGATATTGA TGGATTTGGT	2280
	TATAAGTGGG AAACACATGA AGACATTTTA GATAACGATG CAAAACAACA AATTGCTGAA	2340
25	CAAATTCAG CACAATTTTA AATGTATGCT TAATCTAAGC TCGAACGGGT ATAATATGAT	2400
	TAATATTATA ACAATTGCGT TTGAAGTGAT AACATCAAGG TTAGCAATTT TAAACAAAAT	2460
	GAGTTATCAA GATAACAGAT GTTAAAAGTG AGGAGAATAT AAATGAAAGC ATCACGCATT	2520
30	CTATTCGGTA TCGGTGTTGG CGTAGCAGCT GGTTTTGTAG TTGCACTTCA AGGACGTGAC	2580
	GACAAAAGTG TCAAGAACAA CACGATCGAT CGTACTGCCC CTACTGGTTC AAAATCAGAA	2640
	CTACAACGTG AATTTGAAAC GATTAAACAA AGTTTAAATG ACATTTTAAA CTATGGTGTT	2700
35	CAAATTAATA ACGAAAAGTG GGAATTTGGT AGTTCAATTG GTGGTGAAAT TAAGTCATTA	2760
	CTTGGAAGT TCAAATCTGA CATTAAATCCT AATATTGAAC GTTTACAGTC ACACATCGAA	2820
40	AATTTACAAA ATCGTGGCGA GGATATTGGA AACGAAATTT CTAAGTAGCA GGTTACGTTT	2880
	TCGATCACAA CTATTTTAT TAGTAACAGC ATATTTATTT TTTAAAATTA AATGCCAAAT	2940
	AAACGAGATG ACATTAGAAA TTAGATATTT CTGTGCATCT CTTTTTTAAA ACTCAAATGA	3000
45	ACTTATGTTT ACAAATTATA GGAAGACATT GTTTGTAGTG ATTTTCGCTT AAATCATATT	3060
	TATGAATTGA TTGAAAACAT TGCTTAGGAT TCATTGTGTT ATCCTTGAC TTTGATTACG	3120
	CTTTACTTAA ATCATTATCG ACAAACAACA TACTTATATT TTCATTGAGC CGAACCTTAT	3180
50	ATACACATTA CATATACCTT ACTTGCACAA ATTATTAATC TGGTGTTTAT TATAATTACA	3240
	TATCACTATA TTTTTCAGCAT TTGTATAACT TAGTTGGTCA AAAGATGCTT TTGCATATGC	3300
55		

	TTTCATAAGT GATGCTTTAT TAGCAAGAAT ATGTGTTGCG AGAAATTTGT TCTGCATTCT	3420
	ACTTCTACGC TAGTCAATCA GACAATTTTA CCAATCCCCA CTTTCGCGTT TCAAATCAAA	3480
5	CAATACGTCG CTCCTTTCTT CTTATATAAC AATTCTTCTA ACATGATATG TTACTATTGA	3540
	ATTACTGAAC CTGAGTTAGT TATAATCTAA CTTATATTGA AAAGAGATGA GGCCTAAGAT	3600
	ATGTTTTTAT GTAAAAGACA AATTGATATC AATGCACGAT TTGGTTTGCC TAGAATTGCA	3660
10	TTTATGAGTG CAGTTGCAAC CATCATTATG TTTTATAGTTA GTTATGAAGT AATGTATTTT	3720
	TTATCTAATA CGCCATTATC AGATAGACAT TTTCTCATCT TTTTATTACT TGTATTTATG	3780
	ACGTATCCAT TACATAAAAG TATACATTTA TTATTTTCT TACCATATAG AAAATCGTTT	3840
15	AAAGTTCATA AGTTAACTAA AAGAAAATGG CTTATATTCT ATAATACCTA CGTCAATCAA	3900
	CCTGTACACA AATTTTATTT TTGCATTAAAC TTAATATTGC CGTTAATTAT CTTATCTGCA	3960
	ATGTTTCGTTT ATCTAACAAT TTCAATCCCG CAATATGGAC ATTATTTTAT GTTCTTATTG	4020
20	GCATTGAATT TCGGTATTTT CATTACAGAT TTATTATATT TAAAAATAAT TATATTTTCT	4080
	AATTATGGAC AATATATAGA AGAACATAGT ACAGGTATTA ATATTTTGAA AAAAATTAAA	4140
25	AATCCATATC ATTTATAACA AAATAATTAT AGCAAGGTGT TATTATTTGT TTTTAGGCTA	4200
	TGTAATAgcT tACAATCAAA TGTATATAGA CCTTGTTTTT TTATTTTCAT CAATTTCTAC	4260
	CCCTAAACCT AATGCTCTAG TCTGATGTCA TGGGTATTG ATTGGTGATA ATATAAACT	4320
30	ATGTTATATT CACGATGATT AACTTACAAA GGAGTTTCAA CTATGAAGAT GATAAACAAA	4380
	TTAATCGTTC CGGTAACAGC TAGTGCTTTA TTATTAGGCG CTTGTGGCgC TAGTGCCACA	4440
	GA CTCTAAAG AAAATACATT AATTTCTTCT AAAGCTGGAG ACGTAACAGT TGCAGATACA	4500
35	ATGAAAAAAA TCGGTAAAGA TCAAATTGCA AATGCATCAT TTACTGAAAT GTTAAATAAA	4560
	ATTTTAGCTG ATAAATATAA AAATAAAGTT AATGATAAGA AGATTGACGA ACAAATTGAA	4620
	AAAATGCAAA AGCAATACGG CGGTAAAGAT AAATTGAAA AGGCCCTTCA ACAGCAAGGT	4680
40	TTAACAGCCG ATAAATATAA AGAAAATTTA CGTACTGCTG CTTATCATAA AGAATTACTA	4740
	TCAGATAAAA TTAAATCTC TGATTCTGAA ATTAAAGAAG ACAGCArGAA AGCTTCACAC	4800
45	ATTTTAATTA AAGTTAAATC TAAGAAAAGC GACmAGAAG GCTTAGATGA TAAAGAAGCG	4860
	AAACAAAAAG CTGAAGAAAT TCAAAAAGAA GTTTCAAAAG ATCCAAGTAA ATTTGGTGAA	4920
	ATCGCTAAAA AAGAATCAAT GGATACTGGT TCAGCTAAAA AAGATGGCGA ATTAGGTTAT	4980
50	GTTCCTAAAG GACAACTGA TAAAGATTTT GAAAAAGCAC TATTTAAGCT TAAAGATGGT	5040
	GAAGTATCAG AGGTTGTTAA ATCAAGCTTT GGATATCATA TTATTAAAGC TGATAAACCA	5100

55



AAAAATCCAA AATTATTGAC TGATGCATAC AAAGATCTAT TAAAAGAATA CGATGTTGAC 5220  
 TTAAAGATC GTGATATTAA ATCAGTTGTC GAAGATAAAA TCTTAAACCC TGAAAAACTT 5280  
 5 AAACAAGGTG GCGCACAAGG CGGACAATCC GGCATGAGCC AATAACACAA AACCGAGCGA 5340  
 CCGTGGTTCA AAAATCATAC CACGGCCGCT CGGTTTTTTC GCATTAAAAA TCGGACAGAT 5400  
 GAGCTCATGT TTCAGTATAC TCATCTGTCC GATATCTTTT AATTCCTAAT CGAGTGATTG 5460  
 10 AGGATTGTAG AATCTACGAT TTTCAAGACC AAATATTTTA TCTGTAAACT GACCCCTTGTC 5520  
 AGTTTTTTTA TATGCCTTTT CAAACATATT CATTCTAGCA TCGATATTAT CGATATAGCA 5580  
 TAAAATTTCT GCTTCTTTTA AGTATGGCAG TTTTGGAGAA CCATACTCTA ACTTACCATG 5640  
 15 ATGAGATAAA ATCATATGTC TTAACAACAT GATTTCTTCT CCTTCAATGT TCAATTCACG 5700  
 AGCTGCTTCA ACTACTTCAT CACTCGCAAT CGAGATGTGT CCTAATAAGT TACCTTCGAC 5760  
 20 TGTATACGAC GTCGCAACAG GACCACTCAA TTCTCTAACT TTACCAATAT CATGCAAAAT 5820  
 AATACCACTA TATAACAAAC TTTTGTTTAA CAATGGATAA ATGTCAAAA TTGATTTTGC 5880  
 AATACGTAAC ATCGTTAATA CATGATAGCT TAAGCCACTC GCAAAGTTAT GaTGATGAGA 5940  
 25 ACTAGCAGCT GGATATGTGT AAAATCGTTC TTGATATTTT TTCAATAAAT GACGTGTGAT 6000  
 ACGTTGTAAA TTAGCATTTT CAATATCTAG CAAATAATGA GAAATCTCTT CTTGTATTTT 6060  
 TGCCGGTGAT AAAGGTGCAC CATCTACAAA TTGTTCTGTT TTTAATTGAT CTTCAGTTGT 6120  
 30 CGCTAGTCTA ATTTGGTTGA CTTTCATCTG TTTATTTCCG CGATAGTTTA TGATGTCACC 6180  
 TTTAACATGT ACAATTTCTT CAGGCTTGAT TGTGCCATA TCATTTTGTG TAGCCGTCCA 6240  
 AAATTTGCT TCAATTTTAC CACTTTTATC TTGCAAATGT AATGTCATAT AATCTTTACC 6300  
 35 TTGTGCTGTT ACACCTGTG TAGCTTTATG CACTAAGAAA AAGTGATCAA CTGAATCTCC 6360  
 GGGATTTAGA TTCTCTATAT TTCTCATCGT TTCCCGCCTT CCTCTATTTT GTTTAATGTA 6420  
 ATCACTTCTT TTGATGGAAC AATATTATCT TTTACACATG TAAAGTATAG TACTTGATAG 6480  
 40 TGTCTGATA ATGATCGTAA ATAATTCAAC ATTTTTTCAG TACGTTTTTT ATCAAAATGA 6540  
 ACAAATGCAT CATCAACAAT TAATGGGAAC GGATAATATG GTCTTAGTAC CTTAATTAAA 6600  
 45 CTGATACGTA AAGCTACATA AAGTAATTCT TTTGTAGATT GACTTAGTTC AACAGGATCA 6660  
 TATAATTGAC CATTAAACATG TTTAACCGTA ATTGAATCTT CATTATAGTT AATCATCGTA 6720  
 TATCTGCCAT CTGTAAATG CTTCAATATT TCTACCGCTT CATTATAAAC TTGAGGCAAA 6780  
 50 CGTTTATCTT TAATTTGTTT AATGTGTTCA TCAACTAAAC TTTGTAAATA ACTTAAACTT 6840  
 GCCCAATCTT TTGCGATATC ATTAAGTTGA TTTTAAAGAC TGTGATATTC ATGTCTTAAA 6900

55

GCTTGCATTT CAAGATATTG CTCATTATAT TCGTCAACTT GAGTAGCCAA TAAATGATCT 7020  
 TCTTCTTCAA GTTGTGCAGT TGTTTTTTCA CTAAACTAG AACTTAATTC ATAAGAATAG 7080  
 5 TTTTGGTTCT CAAGATATTT AGTTAAATCA TTAAAACGAC TCAAATTACT AGTATAAGTT 7140  
 TGGTAATCTT CATGATGTTG GTAAAAATCT TCTTCAGTAC CAACATTGAT AAAATCGAAT 7200  
 AGTGCTGTAA TTTCTTTATT ATTTTCTTCT AATTGAGCAT TTAAATGATT TAATTCATTT 7260  
 10 GTAAACAAGTT TGGTATTTTC AGCATTAATA CGCCATTTTT CATTCGTGTC TTCAGCTGAT 7320  
 TTCAACCATT GTtGCACATC GTGGAATAAA GATAATTTGT TGAAATAAAC AAATTGTGAT 7380  
 TTTGTAACAG CTTCAGCATG ATTGTAGAAT GTATCTAATT CTTGAACCAA TTGCTGGCGT 7440  
 15 TGTGTATTTA AATCACTGAT ATGTTGATCT AATGCTTTAA TATTCGCCAT TGTAGAAATA 7500  
 CTATCAACAA TTAAATCATT TGAAATTTTA GATGATAAGT ATAATTCATC CTTAACGTTT 7560  
 TCAACTGTCTG ATTGTAATTC ATCATGACGC CCTTTCGCAT CATTTAAACG ACCTTCAATA 7620  
 20 TACTGACGTT TCTCTTCTAA AATATCTTTA TTTTCAAAG CTTGTTGCCA GTGATCACGA 7680  
 ATGCGATATT GCTCATCAAG ATCAAAATCT AAGTCATAAT TTTCATCTAA AATGGCTAGT 7740  
 25 TGTGCTTTAA TTTCTTCGAT TTCATCTGTG ATGGCCTCGC TATAATCTAC TTCTTTTGAT 7800  
 TTAGACATGA TGATACCGAT AACAAATACT AAAGTTAATA CTGCGAAAAT AATACCAAAC 7860  
 AACATGTTGT TTGAAATAAA TGAGAAGGCA GTTAAACCAA TACCTACTAA TGTTAAAAGr 7920  
 30 ATAAACGTTG TTCGkaACAA TTTTGTGACGT TTTTGtTTT CTT 7963

## (2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 3958 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

ATATTGTCTT TACAATAGTT TGCTATGGAG GTAATTAACC AATAGGAGGA ATTTATAATG 60  
 45 GCAGTAATTT CAATGAAACA ATTACTAGAA GCGGGTGTTT mCttCGGTCA CCAAACACGT 120  
 CGTTGGAACC CAAAAATGAA AAAATATATC TTTACTGAGA GAAATGGTAT TTATATCATC 180  
 GACTTACAAA AAACAGTGAA AAAAGTAGAC GAGGCATACA ACTTCTTGAA ACAAGTTTCA 240  
 50 GAAGaTGGTG GACAAGTCTT ATTCGTAGGA nCTAAAAAAC AAGCACAAGA ATCAGTTAAA 300  
 TCTGAAGCAG AACGTGCTGG TCAATTCTAC ATTAACCAAA GATGGTTAGG TGGATTATTA 360

55

	GAAGATGGTT TATTCGAAGT ATTACCTAAA AAAGAAGTAG TAGAACTTAA AAAAGAATAC	480
	GACCGTTTAA TCAAATTCCT AGGCGGAATT CGTGATATGA AATCAATGCC TCAAGCATT	540
5	TTCGTAGTTG ACCCACGTAA AGAGCGTAAT GCAATTGCTG AAGCTCGTAA ATTAAATATT	600
	CCTATCGTAG GTATCGTTGA CACTAACTGT GATCCTGACG AAATTGACTA CGTTATCCCA	660
	GCAAACGACG ATGCTATCCG TCGGTTTAAA TTATTAAGTG CTAAAATGGC AGATGCAATC	720
10	TTAGAAGGTC AACAAGGCGT TTCTAATGAA GAAGTAGCTG CAGAACAAAA CATCGATTTA	780
	GATGAAAAAG AAAAATCAGA AGAAACAGAA GCAACTGAAG AATAATCAAC TGTTGAATCT	840
15	GACTTAGATA TAGTTTAAAT GGGTGATAAG ATATTAATGC TTATCACCTT TTTTAAAAAG	900
	AAAATCGAGG CAAATTACAA ATATTCAATT AGAGTATTGG CAATCTTGCC TATAATAATG	960
	CTAAAATCAT AATATATAAN ATGATAACTT ATTGGAGGAA TAATGAATGG CAACTATTTT	1020
20	AGCAAACTT GTTAAAGAAT TACGTGAAAA AACTGGCGCG GGTATGATGG ATTGTAAAAA	1080
	AGCGCTAACT GAAACTGATG GTGACATCGA TAAAGCGATT GACTACCTAC GTGAAAAAGG	1140
	TATTGCTAAA GCAGCTAAAA AAGCAGACCG TATTGCGGCT GAAGGTTTAG TACATGTAGA	1200
25	AACTAAAGGT AACGACGCAc TATCGTTGAA ATCAACTCTG AAACAGACTT TGTTGCTCGT	1260
	AACGAAGGTT TCCAAGAGTT AGTTAAAGAA ATCGCTAATC AAGTATTAGA TACAAAAGCT	1320
	GAAACTGTTG AAGCTTTAAT GGAAACAACT TTACCAAATG GTAAATCAGT TGATGAAAGA	1380
30	ATTAAAGAAG CAATTTCAAC AATCGGTGAA AAATTAAGTG TTCGTCGTTT TGCTATCAGA	1440
	ACTAAACTG ATAACGATGC TTTCCGCGCT TACTTACACA TGGGTGGACG CATTGGTGTA	1500
	TTAACAGTTG TTGAAGGTTT AACTGACGAA GAAGCAGCAA GAGACGTTGC TATGCATATC	1560
35	GCTGCAATCA ACCCTAAATA TGTTTCTTCT GAACAAGTTA GCGAAGAAGA AATCAACCAC	1620
	GAAAGAGAAG TTTTAAAAACA ACAAGCATT	1680
	AATGAAGGTA AACCAGAAAA CATCGTTGAA	1740
40	AAAATGGTGG AAGGACGTTT ACGTAAATAC TTACAAGAAA TTTGTGCTGT AGATCAAGmT	1800
	TCGTTAAAAA CCCTGATGTA ACAGTTGAAG CTTTCTTAAA AACAAAAGGT GGAAACTTG	1860
	TTGACTTCGT ACGCTATGAA GTAGGCGAAG GTATGGAAAA ACGCGAAGAA AACTTTGCGG	1920
45	ATGAAGTTAA AGGACAAATG AAATAATCTG TCATAAAGTA AAACAAGGAA GAAGACACCT	1980
	TTAATGTTGC TTTATTAAAA TGTAAATCAT TCTAATAAAA CGACAACTGT GTCTTCTTTA	2040
	CTTGATATG TTACATATAT TCACGATAGA GAGGATAAGA AAATGGCTCA AATTTCTAAA	2100
50	TATAAACGTG TAGTTTGTAA ACTAAGTGGT GAAGCGTTAG CTGGAGAAAA AGGATTTGGC	2160
	ATAAATCCAG TAATTATTAA AAGTGTGCT GAGCAAGTGG CTGAAGTTGC TAAAATGGAC	2160

55

	TTAGGTATGG	ACCGTGGAAC	TGCTGATTAC	ATGGGTATGC	TTGCAACTGT	AATGAATGCC	2280
	TTAGCATTAC	AAGATAGTTT	AGAACAATTG	GATTGTGATA	CACGAGTATT	AACATCTATT	2340
5	GAAATGAAGC	AAGTGGCTGA	ACCTTATATT	CGTCGTCGTG	CAATTAGACA	CTTAGAAAAG	2400
	AAACGCGTAG	TTATTTTTGC	TGCAGGTATT	GGAAACCCAT	ACTTCTCTAC	AGATACTACA	2460
	GCGGCATTAC	GTGCTGCAGA	AGTTGAAGCA	GATGTTATTT	TAATGGGCAA	AAATAATGTA	2520
10	GATGGTGTAT	ATTCTGCAGA	TCCTAAAGTA	AACAAAGATG	CGGTAAAATA	TGAACATTTA	2580
	ACGCATATTC	AAATGCTTCA	AGAAGGTTTA	CAAGTAATGG	ATTCAACAGC	ATCCTCATT	2640
	TGTATGGATA	ATAACATTCC	GTTAACTGTT	TTCTCTATTA	TGGAAGAAGG	AAATATTAAA	2700
15	CGTGCTGTTA	TGGGTGAAAA	GATAGGTACG	TTAATTACAA	AATAAATTTA	GAGGTGTAAA	2760
	ATAATGAGTG	ACATTATTAA	TGAAACTAAA	TCAAGAATGC	AAAAATCAAT	CGAAAGCTTA	2820
20	TCACGTGAAT	TAGCTAACAT	CAGTGCAGGA	AGAGCTAATT	CAAATTTATT	AAACGGCGTA	2880
	ACAGTTGATT	ACTATGGTGC	ACCAACACCT	GTACAACAAT	TAGCAAGCAT	CAATGTTCCA	2940
	GAAGCACGTT	TACTTGTTAT	TTCTCCATAC	GACAAAACCT	CTGTAGCTGA	CATCGAAAAA	3000
25	GCGATAATAG	CAGCTAACTT	AGGTGTTAAC	CCAACAAGTG	ATGGTGAAGT	GATACGTATT	3060
	GCTGTACCTG	CCTTAACAGA	AGAACGTAGA	AAAGAGCGCG	TTAAAGATGT	TAAGAAAATT	3120
	GGTGAAGAAG	CTAAAGTATC	TGTTTCGAAAT	ATTCGTCGTG	ATATGAATGA	TCAGTTGAAA	3180
30	AAAGATGAAA	AAAATGGCGA	CATTACTGAA	GATGAGTTGA	GAAGTGGCAC	TGAAGATGTT	3240
	CAGAAAGCAA	CAGACAATTC	AATAAAAGAA	ATTGATCAAA	TGATTGCTGA	TAAAGAAAAA	3300
	GATATTATGT	CAGTATAAAA	CTAATATACA	ATGACATATT	AAAATGCCAG	TATTAAACGA	3360
35	TAATGTAACA	TTTAAAATGG	GCATGTTTAA	TTAAATCAAA	GATGCATGTG	ATAATTTAAA	3420
	TTCAATGAATGA	GCATAAAAAT	GGTGTTTTAA	CAAGTTAATT	AAACATATAC	TTTATAAATA	3480
	ATAGGCATTA	GGTATATTGC	TATAATAAAG	TTATGTAATT	TTTAACCTCA	GTATGTATGT	3540
40	CACATTTCTG	GTGTAAACTG	TACCGAGTCA	GACTTTGGTA	CAGTTTTTTT	ATTTGCTTAT	3600
	TCAATGCATT	AAATGAGTAT	GATAAAATGA	TAATGATTGT	TTAGTAACTT	ATACTATATG	3660
45	ACAGAGATGA	TCAGGCTCGG	AGGAAAGACC	ATGTTTAAAA	AGCTAATAAA	TAAAAAGAAC	3720
	ACTATAAATA	ATTATAATGA	AGAATTAGAC	TCGTCTAATA	TACCTGAACA	TATCGCTATT	3780
	ATTATGGATG	GTAATGGGCG	ATGGGCTAAG	AAGCGAAAAA	TGCCTAGAAT	TAAAGGTCAT	3840
50	TACGAAGTAT	GCAACAATA	AAAAAAATTA	CTAGGGTAGC	TAGTGATATT	GGTGTTAAGT	3900
	ACTTAACTTT	ATACGCCTTT	TCCACTGAAA	ATTGGTCAAG	ACCTGAAAGT	GAAGTAAA	3958

55

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5333 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

10 ATTAAAACAA CTTAATATAC CTATTTATGG TGGTCCTTTA GCATTAGGTT TAATCCGTAA 60  
 TAAACTTGAA GAACATCATT TATTACGTAC TGCTAAACTA AATGAAATCA ATGAGGACAG 120  
 15 TGTGATTAAA TCTAAGCACT TTACGATTTT TTTCTACTTA ACTACACATA GTATTCCTGA 180  
 AACTTATGGC GTCATCGTAG ATACACCTGA AGGAAAAGTA GTTCATACCG GTGACTTTAA 240  
 ATTTGATTTT ACACCTGTAG GCAAACCAGC AAACATTGCT AAAATGGCTC AATTAGGCGA 300  
 20 AGAAGGCGTT CTATGTTTAC TTTCAGACTC AACAAATTCA CTTGTGCCTG ATTTTACTTT 360  
 AAGCGAACGT GAAGTTGGTC AAAACGTAGA TAAGATCTTC CGTAATTGTA AAGGTCGTAT 420  
 TATATTTGCT ACCTTCGCTT CTAATATTTA CCGAGTTCAA CAAGCAGTTG AAGCTGCTAT 480  
 25 CAAAAATAAC CGTAAAATTG TTACGTTCCG TCGTTCGATG GAAAACAATA TTAAAATAGG 540  
 TATGGAACCT GGTATATTA AAGCACCACC TGAACATTT ATTGAACCTA ATAAAATTAA 600  
 TACCGTACCG AAGCATGAGT TATTGATACT ATGTACTGGT TCACAAGGTG AACCAATGGC 660  
 30 AGCATTATCT AGAATTGCTA ATGGTACTCA TAAGCAAATT AAAATTATAC CTGAAGATAC 720  
 CGTTGTATTT AGTTCATCAC CTATCCAGG TAATACAAAA AGTATTAACA GAACTATTAA 780  
 TTCCTGTAT AAAGCTGGTG CAGATGTTAT CCATAGCAAG ATTTCTAACA TCCATACTTC 840  
 35 AGGGCATGGT TCTCAAGGTG ATCAACAATT AATGCTTCGA TTAATCAAGC CGAAATATTT 900  
 CTTAECTATT CATGGTGAAT ACCGTATGTT AAAAGCACAT GGTGAGACTG GTGTTGAATG 960  
 40 CGGCGTTGAA GAAGATAATG TCTTCATCTT TGATATTGGA GATGTCTTAG CTTAACACA 1020  
 CGATTCAGCA CGTAAAGCTG GTCGCATTCC ATCTGGTAAT GTACTTGTTG ATGGTAGTGG 1080  
 TATCGGTGAT ATCGGTAATG TTGTAATAAG AGACCGTAAG CTATTATCTG AAGAAGGTTT 1140  
 45 AGTTATCGTT GTTGTTAGTA TTGATTTTAA TACAAATAAA TTACTTTCTG GTCCAGACAT 1200  
 TATTTCTCGA GGATTTGTAT ATATGAGGGA ATCAGGTCAA TTAATTTATG ATGCACAACG 1260  
 CAAATCAAA ACTGATGTTA TTAGTAAGTT AAATCAAAAT AAAGATATTC AATGGCATCA 1320  
 50 GATTAAATCT TCTATCATTG AAACATTACA ACCTTATTTA TTTGAAAAAA CAGCTAGAAA 1380  
 ACCAATGATT TTACCAGTCA TTATGAAGGT AAACGAACAA AAAGAATCAA ACAATAAATA 1440

55

## EP 0 786 519 A2

	GCTTTTCTT TATATATGAT GAGCTTGAGA CATAAATCAA TGTTCATGC TCTACAAAGT	1560
	TATATTGGCA GTAGTTGACT GAACGAAAAT GCGCTTGTA CAAGCTTTTT TCAATTCTAG	1620
5	TCAGGGGCCC CAACATAGAG AATTTGAAA AGAAATTCTA CAGGCAATGC GAGTTGGGGT	1680
	GTGGGCCCCA ACAAAGAGAA ATTGGATTCC CAATTCTAC AGACAATGTA AGTTGGGGTG	1740
	GGACGACGAA ATAAATTTTG AGAAAATATC ATTTCTGTCC CACTCCCGAT TATCTCGTCG	1800
10	CAATATTTTT TTCAAAGCGA TTAAATCAT TATCATGTCC AATCATGATT AAAATATCAC	1860
	CTATTCTAA ATTAATATTT GGATTGGTG AAATGATGAA CTCTTGCCT CGTTTAATTG	1920
	CAATAATGTT AATTCATAT TGTGCTCTTA TATCTAAATC AATGATAGAC TGCCCCGCCA	1980
15	TCTTTTCAGT TGCTTCAAT TCTACAATAG AATGCTCGTC TGCCAACTCA AGATAATCAA	2040
	GTACACTTGC ACTCGCAACA TTATGCGCAA TACGTCTACC CATATCACGC TCAGGGTGCA	2100
20	CAACCGTATC TGCTCCAAT TTATTTAAAA TCTTTCATG ATAATCATT TGTGCTTTAG	2160
	CAGTTACTTT TTTACACCT AACTCTTTTA AAATTAAAGT CGTCAACGTA CTTGATTGAA	2220
	TATTTTCACC AATTGCCACA ATGACATGAT CAAAGTTACG GATACCTAAA CTTTTCATAA	2280
25	CTGCTTCATC TGTAGTGTCT GCAACAACCG CATGAGTAGC GATATCACTA TATTCATTCA	2340
	CTCTATTTTC ATCATGGTCG ATGGCCATTA CATCCATGTC TAATGCATT CACTCACGAA	2400
	CGATACTACC TCCAAAACGA CCTAGACCGA TGAATACATA TTCTTTACCC ATACTCGCCC	2460
30	TCCATTAAAT GATTTTCATC AATTCATTGA AAATATAAAT TTAAAATTAT TATAAATGAG	2520
	TACCCCAACT AAATTATCTA AATGCAGTAA TGCAAGTAAA TGAAAGTTGG GGTATCGTCT	2580
	CAACTTATGA TTTCTTTCCT TCAACATATT CTTTGTGCGA AACAAATAAT CTTAATAATA	2640
35	ATATTAACGA TGGAAGTAAT AAAAGTAAAC CTAAATAAA GACAATCACT AATGTCCAGC	2700
	CCATCTCTGG ATTAACATAT GCATCTGTAA TTTTACAAA CGGATATAAA AGGTATGGCA	2760
	ATTTACTAAT TCCATAGCCA AAGAACGCGA ACATCATTTG TAAAATAACA AATACAAAAG	2820
40	CCTAACCATG TTTTCTTCTA AAGAATGTTA ACAATGAAGC TAATGCAAAG AATAAGAAAC	2880
	TTATACCAA CATCCACCA TAGTCAAAA CAGCTGAATA AAAATGTTCA GAATTTTGAA	2940
45	TGCGTAATGA TAGAAATACG AATAACAAA TGATAATCAT CGGCGGCCCT AAAAAATATG	3000
	GCCATGTCT TGTAAATTA TATGCTGGTT CGTCATTGTC TTTTGTAGCA TAATATGTCA	3060
	AAAATCCTGA TGAAATATAT AAAACTGAAA TAATTGCCAA GAATACTACA GACCAAGCAA	3120
50	ATGGGCTTAA TAATAACTGC ACCCAATCTA GATCGATAAC ATTGTTTCGA ACATTAATAT	3180
	AGCCACCTTC TGTAATAGTT AAAGCAGTAG ATAATGAAGC TGGAATTAAT AATCCACTTA	3240

55

	AACTGTTTCT CAACGATATC ATAATCAGTG CTATTGAACC TGGTATTAAAC AATACCGTGC	3360
	CTAAATATTT GATTGACTCT GGAAAGAAAC CTACGAATCC TACGAAGAAG AAAACAAAGA	3420
5	ATACATTTCGT AACTTCCCAA ACTGGGTTTA AATAACGTGA AATTAAGTGA TTAATTTTCT	3480
	TTTCATCACC AGTTAACTTT GAATGCAATG CGAAGAAACC TGCCCCAAAA TCTATAGAAG	3540
	CAATAATGAT ATAGCAAAAT AAAAACAACC ATAACACTGT TATACCTATA AATGCATAAA	3600
10	TCATTTTTCT ATTTCTCCTC CTTGCTTCTT GGCTAAACGA TTTACATCTT CACACGCCGG	3660
	TTTATTTTTTA AACATACGAA TTAATACGTA TGCACATGTA TACATTAAAA TGATGTACAA	3720
	TATGCCAAAT AAAATTGTAA CGAAGGTTAT TCCGCCTGCT TGTGTTGCTG CTTCTGCCAC	3780
15	GCGCATATAA CCACGAACAA TCCAAGGCTG TCTACCCATC TCTGTTAAGA ACCATCCAAA	3840
	TTCTATAGCT AGCATTGAAG CTGGGCCTGT TAATAATATT CCATAAAGCA TCCATTTATG	3900
	AGTAGAAAAC TTCTAAGCT TTTTAAACAT TAAAGTTAAG ACATAAACAC CTGAAATGAC	3960
20	AAAACATAAA ATTCCCATCG TTACCATTAA ATCAAAGAAA TAATGGACGA TCATAGGCGG	4020
	ATGTAAACTT TTTGGAAAAT CATTTAACCC TTGTACTTTA GTTTTGACAC TATTATCTGC	4080
25	TAAGAACTC AATAGTCCAG GTAATTCAT CGCACCTTTA ACTTGCTGAG TCTTTTCATC	4140
	TAACACACCA AATAATAATA ATTTGGCATG GGAAGATGTA TCGAAATGCC ATTCATAAGC	4200
	TGCTAATTTT TCAGGTTGGA ATTTATGCAA AAATTTTGCA GATAAATCCC CTGCCAACAT	4260
30	AGAAAGTAAT GTTGAAAAGA ATCCAACAT CATAGACATT TTCAAAGCTT TCTTATGGTA	4320
	GACAGTATCT TTAGGTTGAC GATTACGCAA TAATTTAAAA GCTGCTATTG ATGCAATAAC	4380
	AAATGCCATC GTCATACCGG CTGTAGTAAT TACGTGAAAT GATCGAACTA TAAACGATGG	4440
35	GTAAACATC GCTTCTATAG GTTGAACATT GACCATCTTT CCATTCTTCA ACTCAAACC	4500
	TGCAAGCGTA TTCATAAATG AATTCACTGA AGTAATGAAG AATGCTGAGA AAGAGCCACC	4560
	AATAATTACT GGTATACTAA TTAAGAAATG TGTCCATTTA TTTTAAAC GATCCCAAGT	4620
40	ATATAAATAT ATACTTAAGA AAATAGCTTC AAAGAAGAAC GCAAATGTTT CCATAAATAA	4680
	TGGAAGTGCA ATAACGTGTC CACCCATTTC CATAAATGTA GGCCAAATCA ATGATAATTG	4740
45	AAGTCCTATA ATTGTACCTG TAACAACTCC CACTGCTACA GTAATTGTAT AAGCTTTAGC	4800
	CCATCTTTTG GCCATAGCTA TATATTGAAG ATCATTTTGT CGAATACCTA AAAATTCTGC	4860
	AATTGCGAAC ATTAAAGGCA TACCAACACC AATCGTTGCA AAAATGATAT GAACTGCTAA	4920
50	AGTCATAGCT GTCAAAAACC GACTGATTTT AACTGTATCC ATTTAAAAAC ATCACCTTTT	4980
	TCTTTTTTTG ATGACAACAC AATGAACTTA ATTATAATTG CTATAATGTG TATTTTTTAA	5040

55

GAATTTCAAT GTATAATTGT GTATATTACA TTAGAATAAA GCACGAAGGA GCATGATACA 5160  
 TGTCAGAAAT AATCGTTTAT ACGCAGAATG ATTGTCCACC TTGTACATTT GTAAAAAATT 5220  
 5 ATCTAAATGA GCATCACATT GATTTTGAAG AGAGAAATAT CAACAATCAA CAATATCGAA 5280  
 ACGAAATGAT AGATTTTGAT GCTTTTTCOA CTCGGTTTAT TTTGTTGAAT GGC 5333

(2) INFORMATION FOR SEQ ID NO: 171:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11126 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

20 ATACGTGACC CTTTATCCGA AAATTCTTT TCATATTCTG TTAATATATT ACTGCCATCG 60  
 TCTTCTTGAT GTAAATTTAG ATTTATTTTT GTAAAATACA TTCAAATTG AGACATACTT 120  
 TCTAACTGT AGGCAAATAG TCCTCTGTTA TCAGTTTAA AATGTAAATC TCCTTCATCA 180  
 25 TTTAAGATTT GTTGATACAA CGCTAAAAAC GTATGATACG TTAACGTCG TTTTGATGA 240  
 CGATTTTTTG GCCATGGATC TGAAAAGTTC AAATAAATAC GCGAACTTC GCCGTCTTTA 300  
 AAATATTCAT TTAATTCAT GCGTCATTA CAAATAATCT TTAATTTGT TAAACCCATC 360  
 30 TCTTTAACTT TATCCAATAC TTTATAAACG ATACTTTTCT CACGTTCCAT TGAAATATAG 420  
 TTAATATGAG GATTTTGAGC AGCTAATGTT GTAATAAACT GCCCATACC CGAACCAATT 480  
 TCAATGTGTA TCGGTTGCGT TTTATCAAAC CATTCAGTCA TTTTCCCTGc ATGTTGACCG 540  
 35 TCCATGTCAA CCAATTCAGG ATGATCTTTT AAATAATCTT CAGCCCATGG TTTGTATCGA 600  
 ACTCTCATAT TTTATCTCTC TCTTAAATAA ACATGTTACT ATTCATAACT TCATTTAGGA 660  
 ATTTAAGCCA AGTGTTTATA TCCTTATATC TTTTTTGCTC TTCATACCAT TGAACAAGAC 720  
 40 CTATAGATTG AATTACCGTA TACCATTTC TACGTTTATT TAAATTCAG CTCTCTTGAA 780  
 CACCATATGT TTCAAGCCAT TCAGACCATT GTTGTGTGG AACATAGTTG TAAAGCAGCA 840  
 45 TTCCGATATC AATTGCCGGG TCTGCAATCA TTGCACCTTC CCAATCAACT AAAAATAGTT 900  
 CATCTCGATC GGATAATAAC CAATTATTAT GATTCACATC ACCATGTACA ACAGTGAAAA 960  
 AACGCGAATC TAACTCGGT ATATGCTCTT CTAAATAGGT TAATGATTTT CTCACAATAT 1020  
 50 GATGTGTAA AACTTCTCTT GATAAAGAGG CATTAATTTT ATTAAGCATA ATCTCAGGAG 1080  
 TAATAGGTTT CATTTCCATA CGCTTAAACA TACTTAATAA AGGTCTAGAA TTGTGTATCT 1140

55



	TTTTCCAATG TTGTGCTGTA ACAACCTCGC CTGTTTCTAT GCGTTTCGTC CATACTAATT	1260
	TGGGCACAAT ACCTTCTGCT GATAATGCCG CAATAAATGG ATTTGAATTT CGTTTTAAAA	1320
5	ACAACTTTTG TCCATCTTGT TCAGCCATAT ATGCTTCACC AGATGCACCA CCTGCTGAAT	1380
	CAAGTGTC CACTAATTGA TAAAACTGCT CCAACTCGTC CACCTCACTT TCAATTAGAA	1440
	AATGGCTCTA GAAATAGGTT TTTCAAGAGC CATATATTCT AATTTATAAC ACCATACTGG	1500
10	TACAAATATT ATGTCCAGAT AATTATTGTA AATCCTCAAC CAATGCCTAC ATTACACGAC	1560
	TAAATTTAAA TCGTAATGTC TGTCATTGAC ACCATACATT CTATAGTCAC TTAATTGACA	1620
15	TATAATGTTA CCGTGTCTAA AACTACATGT TTTTGAATCT CTGTAGGCGA TAAACTCTAG	1680
	TTTTCAAAAT AATTGCTATC CCATTTTCAT GGTTAGCATA AATTTATGAA CTGTAACATT	1740
	TACGTACTTA GTAAATATG ATGCACATCA TATTTGTAC TCATAGAAAA TTTTATAATT	1800
20	TTTATCATT ATTTTCAACT GAAAATGAGA AACAAAATGG CACTTTTAC TAATATGTGT	1860
	TTTCTAAACA ACACTTTTAA GCTTCGTTTT AAATTATAAC ATAATTCATT TACGAAAGTT	1920
	GATAAATTTA AGTAATTTAA TCTAAAAATA TGATGAAAGA ATTTTAAATA CTGTGTGACT	1980
25	CTATATACTT TTCAAATCCT TCTTGTAGTT GACGTGTAAT TGGGCCAACT TTACCATCAT	2040
	TAACTGGTTC ACCATCTAAT TTAATAACAG GTGTAACCTC AGCTGAAGTA CTTGAAACAA	2100
	TAACTTCATC TCGTTTTTTC AAGAAATCTA CAGTAAACGT TTCTTCTTTA AATGGGATGT	2160
30	TATAGTCTTC GGCAATTTTT TTAATTACAA TTCGTGTAAT ACCATTAAGA ATATAGTTGT	2220
	TAATCGGATG TGTATAAATC ACACCGTCTT TAATTGCATA AGCATTACTT GAAGATCCTT	2280
	CAGTTACAGT TCACCTCGA TGTTGAATTG CTTCAACTGC ATTATATTTT ACAGCATATT	2340
35	CTTTTGCTAA TACATTCTCC TAATAAGTTC AAGCTTTTAA TGTGCAACG TAACCATCGG	2400
	ATATCTTCAA CGGTAACACC ATTCACACCA TTTTCTAAAT GATCATAAGG ACGATCATAA	2460
40	CTCTTTGTAT AAGCAACAAT TGCTGGTTCT ACTTCAGGTG TCGGGAAGCT ATGATTCTTT	2520
	TCAGCTACAC CACGCGTTGC TYGAATATAA ATTGCCCCAG TTTCAATTTG ATTCATATCA	2580
	ACTAATTTAC GAGATAGTTC AATTAATTCT TCTACAGAAT AATTTAAATC TAAACCAATC	2640
45	TCATTGGCAC TACGTWAAAW TCTTTCATAA TGTTCTGTTA CTGTAAATAA CTTACCATTA	2700
	TATACTCGAA TGTATTCATA AATACCATCG CCAAATACGT ATCCTCTGTC GTTGTATGAA	2760
	ACCTTGCTT CACTTGGACT TACAACTCA CCATTTAAAA AAATTTTTTC CATATATTAT	2820
50	TCCTCCACGC ATAATGAATA AATTGCTTCT AAGTAAATAC TAGTTGCGTT AAATAACTGT	2880
	TTTTTAGTGA TATATTCATT TTTCTGATGC ATTAAATCTT CAGAATCACT AACATTGCG	2940
55		

TCAGTCATAT CATTGTGTTG ATTTCTATAT GCAGTAACTA ACTTTTGTAC AAAAGGATCA 3060  
 TTTTATCAA CATAATGTGG TGGTTGGA CT TACCTAATT TCACTTCAAA GCCATATTGT 3120  
 5 TGAATCTCAT TTGCAAAACG ATCCATAGCT TTTTCAAATT CAAATCCTTC TGGGTAGCGT 3180  
 AAGTTGATAC CGAAAAGACC TGC GTTTTCA TTATCATATG TAATAACACC AATGTTAGTT 3240  
 GTCACGTCAC CCATGACATC TGTATGGAAT TTCATTCCCA TCTTTTCACC AAAATCTGAA 3300  
 10 TTAAATAAGT AGCGATTACT AAATGCTACA AACGCTTG TG CATTATTATC AAGATTTAAT 3360  
 GATGCTAAGA ATTTTAGTAA GTAAAGACCC GCATTACAC CGATAGATGG ATCCATACCA 3420  
 TGAACCGCTT TACCTTCAAC TGT TAAACT AGAATGCCAC TATCAACAGT ACTATCACCT 3480  
 15 TGTAATGAT TTTGTTCTAA AAAGTACTCA AAGTCTTGAA TAACATCTGT CATATTTTCT 3540  
 TTAACAAGCA CTCTTGCTTC TGCATGATCA GGTACCATGT TGTAACGTTC ACCAGATTTA 3600  
 20 AAAGTTATTA ATTCATAATC AGGTTTCTCT TGATCTTCAG TAAGTTTATT TTGAACTAAA 3660  
 TCAAATGTTG TAATGCCTTT TTCACCATGA ATACATGGAA ATTCTGCATC TGGTGCAAAA 3720  
 CCTAATGTTG GCATTTCTTC TGT TTTAAA TAGCGATCCG TACATTTCCA ATCAGATTCT 3780  
 25 TCATCCGTAC CAATAATCAT ATGAATACGT TTCTTCCAAT CCACATTCAT ATCTTCTAAT 3840  
 ATCTTAATTG CATAATAAGC AGCAATTGTT GGACCTTTGT CATCAAGTGT ACCTCTAGCT 3900  
 ATGATAGCAT CTTCTGTTAC AACC GGCTCG AACGGATTAC TATCCCATCC ATCACCAGCA 3960  
 30 GGAACAACGT CAACATGACA TAAGATACCT AATACGTCAT TTCCTTTACC TGCCTCAATT 4020  
 CTTCCTGCAA TATGATCCAC ATCATGTGTT GTAAATCCAT CTCTATGTGC AATTTCATAC 4080  
 ATGTAGTCTA ATGCCTTACG AGGACCTGGA CCAACTGGTG CGTCTTCTGA TGCTTTTGCA 4140  
 35 TCATCTCTCA CACTTTCAAT TGCTAATAAT CCTTTTAAGT CATTAATGAT TTGATCTTCG 4200  
 TATTGTTGAA CTTTTTCTTT CCACATTCGA AATCGACTTC CTTTTTTCTA TAAGTTAAAT 4260  
 TCTATTTTAC ATGAAAAGAT ATAAAACTA CAATAAGATG TCAGAAAATA ATAAAAAGGA 4320  
 40 ACAAACGAT GCTATTGATA TGACACAAAT CATAAATAGC TGCTTTGTTC CTTTTTTAAT 4380  
 TTATATATTT AAAATACACA TATTCAAGAG CTCGAGATAT AAGTCAATGT ACTAGGCACA 4440  
 45 CAATTTAATA TTGACAGTAA TTAACCGAAC GAAATGCGC CCCGGGGCCC CAACATAGAG 4500  
 AATTTGAAA AGAAATTCTA CAGACAATGC AAGTTGGCGG GGCCCCAACA TAGAAGCTGG 4560  
 CCAATAGTTA GCTTTCAATA ATGTGCAAGT TGGGGTAAGG GCCCCAACAC AGAAGCTGGC 4620  
 50 CAATAGTCAG CTTTCAATAA TGTGCAAGTT GGGGTAAGG CCCCAACACA GAGAATTTCTG 4680  
 AAAAGAAATT CTACAGACAA TGCAAGTTGG CGGGGCCCA ACACAGAAGC TGGCCAATAG 4740

55

	TAAAGAAATA CGTTTTCTTT AGATATTAGT ATTTCTTATG AATGAGTTTC ACGCATGTAT	4860
	TCTTCTTTCT ATATGCATAT TAGCTATGAC TAACGATAAA GAACCTGAAA CACTAATAAA	4920
5	TGTCCTATAG TTTACAATAT TATATTGGCA GTAGTTGACT GAATGAAAAT ACGCTTGTA	4980
	CAAGCTTTTT TCAATTCTAG TCAACCTTGC CGGGGTGGGA CGACGAAATA AATTTTGCTA	5040
	AAATATGATT TCTGTCCAC TCCCTTATCA TTTCTGTCT ACTCACATCT TATTCTTTAT	5100
10	CAGATAATGC ATTTTATTTC TTTTTAAAT CTTCTTCAGT GACGATACGT AAATTATTAT	5160
	TTGGTGTGCG CCACCTTCAT CATCAAATTT ACCTTTTCA ATACTTTCGT CAGTCTTATT	5220
	GTCATATTCG GTAAATTTTG ATTTTCTTC TTTGAAAAAT GCTTTTGAT TATTTTTTAA	5280
15	TCTATTAGCA TATTCTTTCG GATTTGTTTT TACTTCTTA ATTGTTTCAT TAGCAATTGT	5340
	TCCTAATTGC GTCGCTTAT CCTTAGCATT ATCTTTATAG CTTTGAGGAT CTTGTTTATA	5400
20	TTTATTATAT TCCTGCTTTC AGCTTGTAC GACTATCTT ACGTGTAACA AGTACAGCTG	5460
	CTACAGCGCC ACCTATACCT AAAATCGCTT TAAATAAAT ACCTTTTGCC ATATCAATCG	5520
	TCTCCCTTTT ATTTATAATT TAATTTGTCA AAATCATTTT CAGTTAATAA ACGATATTCT	5580
25	CCTGAATCTA AATGCTGTC CAATTCTAAA TCAGCAATTT TGATACGCTT TAAATGTAAT	5640
	ACCTCATTTT GAATGCTATG AAACATTTCG TTAACCTGAT GATATTTTCC TTCATAAATT	5700
	GTTACGTGTG ACGTTTGATT ATCAATATAA GTTAATATTG CAGGCTTAAC CTTGCCATCA	5760
30	GTCAGTGTA CACCTCTTT AAAAGCTTGA ATGTCGTCTT CAGTGATAGG ATTTGCTGAA	5820
	ATAACTTCAT ATTTTTTAGA AACATGTTTG TTTGACTCA TTAATTCATG ATTAAATCA	5880
	CCATCATTCG TTATCAATAA AAGCCCTTCT GTATCTTTAT CAAGACGACC AACCAGAAAA	5940
35	ATATTTAGAT GTTGGTATTC AGGTATTAAA TCAATAACGG TTTTGAATG ATGATCTTCA	6000
	GTTGCTGATA TATAACCTTT TGGCTTATTT AACATAATAT AGACATTTTC AATGTATTCT	6060
40	ATTAAATCTC CACGAACGT TATCTTATCG TTTCTGGT CTATATGTGT TTTTGGTGAT	6120
	TTAATTACTT GTTCGTTGAC ATTTACAAGG CCTTTTTTAA GTAAGTGTG GACCTCATTA	6180
	CGTGTACCGA CGCCCATATT TGCTAAAAAT TTATCTATTC TCATCGTAAA AACCTAACTC	6240
45	TACGTCTTAA TTTTTCAGGA ATTTACCTA AGAATTCGTC CGCAAGACGC GTTTTAATTG	6300
	TGATTGTACC GTAAATTAGA ATACCTACTG TAACACCTAA AATAATAATG ATTAAGTAAC	6360
	CAAGTTTAGT AGGTTCTAAG AATAGATTTG CAAGGAAAAA TACTAATTCT ACACCTAGCA	6420
50	TCATAATAAA TGAATACAAG AATATTTTGG CAAAATGAAT CCAACTATAG CTGAATTTAA	6480
	ACTTCGCATA TTTTTTAAGA ATATAGAAAT TACATCCAAT TGCAAATAAT AATGCGATAC	6540
55		

	ACTTGATAAC TACAGAAGCT AAAATAACAT AAAGTGTAA TTTCTGTTA TCTATACCTT	6660
	GTAACATTGA TGCCGTTACA CTTAATAGTG AAATTAGTAT TGCTACAGGC GCATAATAGA	6720
5	ATAATAAGCG ACTACCATCA TGGTTAGGGT CATGACCTAA AACAAATTGA TCGTAACCAT	6780
	AGAAAAGTGT GAATAATGGT TGTGCCAAGG CCATAATTCC AATACTAGCT GGAACAGTTA	6840
	TAAACATTAA TACACCAATA GATGTTCTAA TTTGATGATG CATTTTCATGT AAGCGACCTT	6900
10	CTGCAAATGT TTTTGTAAATA TAAGGAATTA AACTCACTGC AAAACCAGCA CTTAATGATG	6960
	TCGGAATCAT TACAATTTTA TTAGTTGACA TATTTAGCAT ATTAAAGAAT ATATCTTGTA	7020
	ACTGTGAAGG TATACCAACT AAAGATAAAG CACCGTTATG TGTAATTTGA TCTACTAAGT	7080
15	TAAATAATGG ATAATTCAAA CTTACAATAA CGAACGGTAT ACTATAAGCA ATAATTTCTT	7140
	TATACATCTT GCCATATGAC ACATCTATAT CTGTGTAATC AGATTCGACC ATACGATCAA	7200
20	TATTATGCTT ACGCTTTCTC CAGTAATACC AGAGTGTGA TATACCAATA ATCGCACCAA	7260
	CTGCTGCTGC AAAAGTAGCA ATACCATTGG CTAATAAAAT AGAGCCATCA AAGACATTTA	7320
	GTACTAAATA ACTTCCGATT AATATGAAAA TCACGCGTGC AATTTGCTCA GTTACTTCTG	7380
25	ACACTGCTGT TGGCCCCATA GATTTATAAC CTTGGAATAT CCCTCTCCAT GTCGCTAATA	7440
	CAGGAATAAA GATAACAACC ATACTAATGA TTCTTATAAT CCAAGTAATA TCATCGACTG	7500
	ACCAACCGTT TTTATCATGA ATGTTTCTAG CTAATGTTAA TTCAGAAATA TAAGGTGCTA	7560
30	AGAAATACAG TACCAAGAAA CCTAAAACAC CGTAATACT CATTACAATA AAAGTCGATT	7620
	TATAAAATTT CTGACTTACT TTATATGCCC CAATAGCATT ATATTTGCA ACATATTTG	7680
	AAGCTGCTAA TGGTACACCT GCTGTCGCAA CTGCAATTGC AATATTATAT GGTGCATAAG	7740
35	CGTATGTGAA CGGCGCCATA TTTTCTTGTC CACCAATTAA ATAGTTGAAT GGAATGATAA	7800
	AAAGTACGCC CAATACCTTG GTAATTAATA TACTAATGGT AATTAAAAAG GTTCCACGCA	7860
40	CCATTTCTTT ACTTTCACTC ATTACGAATC TCCCTATCTC ATGTTTATTA AAGTTTGTGTA	7920
	AACTAAAAGC TGTTTCTCTG TAAAATCATT TTTCATTATT ATGAATATAT CACAAAACCTT	7980
	TATTTTCATTG TCGTATATC AATGAATTAT CATAACAAAA TTATCAACAC ATTGTCATTG	8040
45	AATACTAGAT TTTGATTAGA ATATTACGAA ATTTTCATATA AACATTATAC TACTATTTGA	8100
	GATGAACATC GCATAACAGT AGAAAAATCA TTCTTATCAT ACACATACAT CTTCATTTTTT	8160
	TATGAAGTTC ACATTATAAA TATATTCAAC ATAATTGTCA TCTCATAACA CAAGAGATAT	8220
50	AGCAAAGTTT AAAAAAGTAC TATAAAATAG CAATTGAATG TCCAGTAACA AATTTGGAGG	8280
	AAGCGTATAT GTATCAAACA ATTATTATCG GAGGCGGACC TAGCGGCTTA ATGGCGGCAG	8340
55		

	GTAAACTCAA AATATCTGGT GCGGGTAGAT GTAACGTAAC TAATCGATTA CCATATGCTG	8460
	AAATTATTAA GAACATTCCCT GGaAATGGGA AATTTTATA TAGTCCCTTT TCAATTTTGT	8520
5	ATAATGAATC CATCATAGAT TTTTTGAGT CTAGGGGTGT TAAATTAAAA GAAGAAGATC	8580
	ACGGGCGTAT GTTTCCAGTT TCCAACAAAG CACAAGACGT GGTGATACA TTAGTGACAA	8640
	CTATCGAACG CCAACATGTA ACGATTAAAG AAGAAGAAGC TGTTAGTAGA ATCGAAGTTA	8700
10	ATACAGACCA AACTTTCACT GTACATACTC AAAATAATAG TTATGAAAGC CATTGCTAG	8760
	TGATTGCTAC AGGTGGTACA AGTGTCCCTC AAAGTGGTTC AACTGGTGAT GGTTATAAGT	8820
	TCGCACAAGA TTTAGGTCAT ACCATTACTG AGTTATTCCC GACCGAAGTT CCAATTACAT	8880
15	CAGCTGAACC TTTCATCAAA TCCAATCGTC TAAAAGGTTT AAGTTTAAAA GATGTTGAAT	8940
	TGTCAGTACT TAAGAAAAAT GGTAAAAAAC GCATCAGTCA TCAAATGGAT ATGTTATTTA	9000
20	CTCATTTTGG TATCAGTGGT CCAGCTGCAT TAAGATGTAG TCAGTTTGTT TATAAAGAAC	9060
	AAAAAATCA AAAGACACAG CACATTTCTA TGGCAATCGA TGCATTTCTT GAATTAAACC	9120
	ATGAACAATT AAAACAACAC ATCACATCAT TATTATCGGA CACACCAGAT AAAATCATT	9180
25	AAAACAGTTT GCATGGTCTA ATTGAAGAGC GCTACTTACT GTTCATGCTG GAACAAGCAG	9240
	GAATCGATGA AAATACCACA TCACATCACT TATCAAATCA ACAATTGAAC GACTTAGTAA	9300
	ATATGTTTAA AGGGTTTGTA TTAAAGGTGA ACGGGACATT ACCTATAGAT AAGGCATTTG	9360
30	TCACAGGTGG TGGTGTGTCA CTTAAAGAAA TTCAACCTAA AACAATGATG TCTAAATTAG	9420
	TTCCGGGATT ATTTTATGT GGTGAAGTAT TAGATATACA TGGTTATACT GGTGGTTATA	9480
	ATATTACAAG TGCACTCGTA ACAGGACATG TCGCTGGATT ATATGCCGGA CATTACTCAC	9540
35	ATGCATCAAT GGAATAATAG TATAAAATTT GGTTCGATTC TCTTTAGTAG ATCAACTTTT	9600
	TCATTCAAAT AAAAATGACC TTAATATAAC TGAGTCACTA AAAAGTGTCTG TTATATTAAG	9660
40	GTCAITTCGT TAATTATGAT TCTTTTTCGT TTTTAGTACG TCTTCTAGCT AACAAAGCCG	9720
	CACCTGTAAT CAGTGCAAAT TCTTTCAATG GTAAATCCAT TCCTTCAGAA CCTGTATTTG	9780
	GAAGTCTTTT TTCAACTTTG CGCGATTCAT GTGTCTCTTC TTTTAAATA GGCGTACAAA	9840
45	CTTTTGGAGC TGGCTGAATT TCTTTTGGTG ATACTTTTCGT CGCTTCAGCT GGTAATTTAA	9900
	TTGCTAAAAT TTCATCAACA ATGAATTGCG TGTGTTGTTT GATGTCATTT AATGTCGCAT	9960
	CTTCATCAAT CATTCATTG CCATCTGCAA CATATTGATC AATTAATACT TTTACTTTAG	10020
50	CTAATGTTC TGGTGTGCG ATCGCTTTGA ATTTGCGATA TGTTTGTGTA GCAATGTTAT	10080
	CAATTCGCAG TAAGCTATTT TCTTTTTCAG TAATTACTGC TTCTATATCG CTTAATGCAA	10140
55		

CATCCATTG TAATTTTAAA GCAGTTATAG CTTTAAATGC ATCAGCCTTA TTACGATTAC 10260  
 TTACTTTTCG ATAATTTTGC ACTAAAGCAG TGACGCGTGC AAGATCATCA TTAATCGTTT 10320  
 5 TTTCAGCATC TGGCTTTTAA ATAGGATGTA CATCTAAATC ATGTATTGTT TGTAGATTAA 10380  
 ATGATGCTGT TTTATCAACT TGTGCATTGC TACGATCTTG ATCAATTTGT CCAATAGCAG 10440  
 TGTCAATAAT ATTTTGTAAC TGTGCTAATA TACTATTTCT TTCTTCTACC GTTGCTTGAA 10500  
 10 TATTCGCTTC AATTGCTTGT TTTTATCGT TGAATAATGT TGTCAATTGT TCTCGAGCAG 10560  
 ACGCCTTTCT GTTAATAACA GGTTCGATT CACGAATTC GTTTTCTCA TCATGCAATA 10620  
 AATATGCCAC ATCTGCATTA GTCACGAC TAGCAATTTG TTGTTTAGCT TTAATTAAC 10680  
 15 CTTTTCAAC TTGTGCTATT GCAATATTTT GTTCTTCATC TGTCGCTTCG TTATTTGCTT 10740  
 TAATTAAATT AATTTTATTT GTAGCGATAT TTTGAATTTG TTGTAATGCT GTTGCTTTAA 10800  
 CTGTTGTCG TGGTTTAATT TTTGAAATAA TATTTTGAGC ATTTATACTA TCTTGATTAA 10860  
 20 CTTGGGCAGT CTTATCTGCA TGATTGATCT GATCAATAGC CTGATTAAAGT GCTTGTTCTA 10920  
 CTAAATGTTT AGCAGCTAGT CTTTCTTCTT CAGTTGATAA ATCGCTTTGA TCGATTAGTG 10980  
 25 CATTTTGAGC TTCGGCTTTT ACACCAACAG ATTGACGCGC TGCTGGTTTA ACTTGAACCT 11040  
 TAGGTAAAAT CACTTTGATG TTGTCGTTGC CATCAGTCnC AGTnCGATCC ACTTCTGCAT 11100  
 TCGTTTTGTT TTGTGCAATG TCATTT 11126

30 (2) INFORMATION FOR SEQ ID NO: 172:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3660 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

TTGCCCCGCA CGGCGGTGTG nTTCCTAGAA ATAATGAATA TAAAGaGAAA TATATAACAA 60  
 CGATTTTGAA TTATGAACCT GGTGATATCG TTACAATCAA ACGTGTGAGA GATAAGACCG 120  
 45 ATTTGCTAAT ATATTTGTCT AGTAAAGATA TTTCTATTGG TAATGAAGTG GAAATTGTAT 180  
 CGAAAGATGA AATGAATAAA GTAATTATCA TTAAACGTAA TGATAATGTA ATTATTGTCA 240  
 GTTACGAAAA TGCAATGAAC ATGTTTGCTG AAAAATAAAA TAAAGAAGCC ATAAAGATAT 300  
 50 CCATGATTGA ACTGATAAAG ACATATGGAT AATTGCTTTA GGCTTCTTTT TTATTAGTTA 360  
 ATTTATCAAG TGAGTATATT TGAGTAAAAT ATTCACTGCA TAAAGATTGA AGATAATCCA 420

55

## EP 0 786 519 A2

	CTGTGGACTC GGACGCTGGA AAGTCAATTT AGCAATCGTC CAACTAGATT GTAGAACTTC	540
	GCCTAATAAT ACACCTAAAA TATATTGATA ACTCATTGTG ACAAGTAGTT GAATTTCTAC	600
5	TATATTTTCA TCTTTTAATA TAAAATACAA CATGATAGAA ATTAAAGTTA TAACAACAAT	660
	GGGTGAGCCT TTTCTAGATG TTAAAATTAA AAAATAAATA AATATCAATA AATAGGTAAA	720
	TATAAAGAAA CTAGGTATCT GATAATGGCT CGACGCTAAA CCTATCAATA ACATAATAGG	780
10	TGGCATAAAA TAACCACCAA TCGTTGTAAG CCATTGGCCT GCTAGATGTC TAGATTGTGT	840
	AATTGCGAAT CCTTGTGTGA ATGTCTGTTG TCGCTCTCGT GGACTTGTGA CAATGACTAA	900
	ATCTTTTGCA CGGCCACCAG CGAGTTTATT AAACAGTACA TGACCAAATT CATGTGTTAA	960
15	AACAGGGATA TAGTTTAAAA TGACATCTAA ATAGTTCAAA ACAGGCTTAT GTCTATATTG	1020
	ATGAATAGCA ATATAACAAG CTGCAACAAT AACGATAATG TATATATTAA GTTGAATTGT	1080
20	CGTATTAAAA AAGTTTGATA AATAATTCAT TGTAAACCTC ATATAAGATA TTAATTTTAA	1140
	GTTTGCTTAT CACTTATTAT AAATGATATT GGCATCAATA GCGTTAGACT TTAGACTTAC	1200
	CTTAGTTAAA CTAATTTTAA TTTTGTAAAA GGTGAATATG TGTAAAAATA AAGCAAATC	1260
25	ATTTGATAT AAATAGGATG AATATAAATA CTGTTAATAT TGATTACACT AACATAATAA	1320
	TGAAATAAGA TAGGAGATTC CTGTTATGAC TGTGAAGAA AGATCCAATA CAGCCAAAGT	1380
	TGACATTTTA GGGGTCGATT TTGATAATAC AACATGTTG CAAATGGTTG AAAATATTAA	1440
30	AACCTTTTTT GCAAATCAAT CAACGAATAA TCTTTTTATA GTACAGCCA ACCCTGAAAT	1500
	AGTGAATTAC GCGACGACAC ATCAAGCGTA TTTAGAGTTA ATAAATCAAG CGAGCTATAT	1560
	TGTTGCTGAT GGGACAGGAG TAGTCAAAGC TTCGCATCGT TTAAAGCAAC CTCTAGCGCA	1620
35	TCGTATACCT GGTATTGAGT TGATGGATGA ATGTTTGAAA ATTGCTCATG TAAATCATCA	1680
	AAAATATTTT TTGCTAGGGG CAACTAATGA AGTTGTAGAA GCGGCACAAT ATGCATTGCA	1740
40	ACAAAGATAT CCAAACATAT CGTTTGCACA TCATCACGGT TATATTGATT TAGAAGATGA	1800
	GACAGTAGTG AAcGnAnTTA AACTGTTTTAA ACCTGATTAC ATATTTGTAG GTATGGGATT	1860
	CCCTAAACAA GAAGAATGGA TTATGACACA TGAAAACCAA TTTGAATCTA CAGTGATGAT	1920
45	GGGCGTAGGT GGTCTCTTTG AAGTATTTGC TGGGGCTAAA AAGAGAGCGC CTTATATCTT	1980
	TAGAAAATTA AACATTGAAT GGATATATAG AGCATTAAATA GATTGGAAC GTATTGGTAG	2040
	ATTAAAGAGT ATTCCAATAT TTATGTATAA AATAGCCAAA GCaAAAAGAA AAATAAAAAA	2100
50	GGCGAAATAA TCATGATGAC AAAAATAAAA CCGAGGAAAT CCTTAAATGG AGATTCTCGG	2160
	TTTTTTCGGT TTATTTAATA ACGAAGCGGG ACTCATCGAG TTTGTTTCTA AATTCITTTT	2220
55		

	CATCAAGTTC ACCGTAATCT TTAACTTTTC CGCCTTCAAT CCAAGCAATC TTAGTACAAA	2340
	ATTGTCTCAC TTGTCCTAAG TTATGACTAA CGAAAAAGAT GGTTTTGTTT TGCTCTTTAA	2400
5	ACTCGTAAAT TTTATCTAAA CATTTTTGTG CAAAAGTTTG GTCACCTACA GATAAAGCTT	2460
	CGTCAATGAC TAAGATATCT GGATTAAGTG TGATATTAAT TGAAAAACCA AGTTTTGCAC	2520
	GCATACCACT TGAATACTTT TTAAGTGGTT GATAAATAAA CTCACCAAGT TCACTAAATT	2580
10	CAATAATCTT AGGTGTCATC GCTTTAATTT CTTTTCGCTT AAAGCCCATA CATAACATTT	2640
	TAAATTCGAT ATTTTCAATC CCTGTAAGTT GTCCACTCAA GCCAGCACTA ATTGCGATAA	2700
	CGCTGACTTC ACCATTACGA TCCACTTTGC CAACAGTAGG CGACAAAGAA CCGCCAATGA	2760
15	TATTGCTCAA CGTTGATTG CCGGAACCAT TGATGCCAAC AAGCCCTATG ACGTCACCTT	2820
	CATATGCTTT TAAACTAATG TCATCTAAAG CGAAAAATGT TTTGTTTTTA TGTTTGGGAA	2880
	TGAGCGCATC TTTCATACGT TCTTTATTTG TACGATAAAT ACGATATTCT TTTGTTACAT	2940
20	TTTTAATGTT TACCGAAACG TTCATTTGTA GACCTTCCTT ATTCACATTT ATCTAGATTA	3000
	TAATATACTA CTCAACAGTT GTTAAATTTT AAAACCTGTT GTAAAGTGTA TAGAAGATTT	3060
25	TGTTATTATC AGAGTGGGTG TTTTGACACA AAATGTTAAT CATCAATGAT AACAAATGATA	3120
	TTAAAAACT AAACCTATTT CAACTTACAT GATTGTATAC TATAATGTAT TTGTAATAAA	3180
	CTAATATTTT AAAGAACTAG ACAATAATTT TGATAGCATC CATGTATAGT GATAGTATTT	3240
30	ACAACAATTA TTATAATACT ATTTAGTTAA GTAGAGAAAT AGTTAAACAT TTGAAAGTGT	3300
	GGTTTAATGG AATGTCAGCA ATAGGAACAG TTTTAAAGA ACATGTAAAG AACTTTTATT	3360
	TAATTCAAAG ACTGGCTCAG TTTCAAGTTA AAATTATCAA TCATAGTAAC TATTTAGGTG	3420
35	TGGCTTGGGA ATTAATTAAC CCTGTTATGC AAATTATGGT TTAAGTGATG GTTTTTGGAT	3480
	TAGGAATAAG AAGTAATGCA CCAATTCATG GTGTACCTTT TGTTTATTGG TTATTGGTTG	3540
	GTATCAGTAT GTGGTTCTTC ATCAACCAAG GTATTTTAGA AGGTACTAAA GCAATTACAC	3600
40	AAAAGTTTAA TCAAGTATCG AAAATGAACT TCCCGTTATC GATATACCGA CATATATTGT	3660

(2) INFORMATION FOR SEQ ID NO: 173:

- 45 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13868 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

55



EP 0 786 519 A2

	ATTAATCACT TGTGTGTAG AGTCTTGTC GTTTTGGTTA TGATTGTTAG CCATGATATA	120
	CCTCCCTTAC AACACTCGTG GACCAGAAGT TTTCTGATCT CTCACATTAA CTTCTAACTT	180
5	ACGTACTGGC ATTTCTGTGA AATATTCTAC ATTCTTTTAA ATATCCGAAC GAATTGCTTC	240
	AGTTAAAGAT TGAACCTGAA CATTATTTGG TACGAAAAAG TCAGTTTAA TGTGATATA	300
	AGATTTATTT TTTTGTAT ATAGTTTCGC AACTACATTT GGTGTCTTA CTTGATCATA	360
10	TTTTGCAACC GTATCGAATG CCGTCTTTTC AACAGCTTTA CGAGATACGT AAACATGACC	420
	ATCATCGAAG TCTTGTATA ATCCAGGTTT TCGATGCGTA GGTTTGAAGA TACTAAATAC	480
15	TAATATAAGA CCTATTAATA TCAATAGTGC AGCAAGTGAA ATAAGTAATG GTTGGAACCA	540
	TTCAAATTGA AGGAAGTAGT CTTGATATTC AGTTATACGT CCATCTTGA TATACATGAA	600
	TAACAGGAAC CCCACGATT CTACTATTAA TAAGCCAAGG ATAAAGTTTT TAAGTCGTTT	660
20	CACCCCTAAC GACACCTCCT TAGTTAAAGT TAATTTAAAA ACATATTAAA TATGTACCCA	720
	TCAGTTTTTT TCTTAAACAT AATAAATTAA TAACTTTAAA TTTATTTTAA ATATATAAGA	780
	TGAAGTACCA TTTAGTAATA TATCCCTAG TTTTGTAAA TAAAACCTCA TTATTAATTA	840
25	ATTTCGTCA ATATGTTTTG AAGAACGATA TTCTAAAATA TCTGGGTCAC GATGTTTAA	900
	TAAAACCTTA TTAATTTTC TCGGTTTCTC CTCACTCAA GATTTTATA GCGACCATAT	960
	CATCGCTATA ATGACCACGG AAAATGGTAA CGCAGCAATG ATTAATAAAT TTTGAATTGC	1020
30	TTGAGTACCA CTTGTGTAAA TCATGATGAT TGCAAATAAT GCCATAATGA TACCCCACT	1080
	CACTTTGACA AATGACTTCG GATTAATATC ACCACTTGAA CTCAACATAC CTAAACATA	1140
	AGTTGCCGAA TCCGCTGATG TAACAAAGAA AATCATAATA ACAAGTAAAG TAATTAAGCT	1200
35	TAATACAAAA CCTAGCGGAT AATGTTGTAG CGTCGCAAAA GTTGCTGTTT CTGTCGCAGC	1260
	TTTAGCAATA TCGGCAATAT GATTATCTTG TAAGTAAATT GCTGACGCGC CGAATACCGC	1320
40	AAAGAATATA AAGCAAATA ACGCCGGGAC AAAAAGTACA CCTAGAATAA ATTCTTTAAT	1380
	CGTACGTCCT TTTGACACAC GTGCAATAAA TATACCTACA AATGGTGCCC AAGATATCCA	1440
	CCATGCCCAG TAAAGATTG TCCAATTTTG TAACCATTTG AATTTTGTAC CACCTGTCGG	1500
45	AATGCGTAAA CTCATACTAA AGAAATTGTC AATATAATTA CCTAGACCAT TCGTAAATGT	1560
	ATTTAAAATG TATAGCGTTG GCCCAACAAT AAAAAGACCA ATAAGTACTA CAAAAGCAAG	1620
	TAACATGTTG ATATTACTCA ACGTTTGTG ACCTTTATCG ATACCTGACC ATGCTGACCA	1680
50	AGTAAATAAT ATGGTTGCAA TGACAATCAA GATTACTTGC ATCGTGAAGT TACTCGGTAC	1740
	ATTAAATAAA AAATGTAAAC CTTGCTTTAT TTGCAATGCA CCGAAACCTA ATGTTGCAGC	1800
55		

EP 0 786 519 A2

	CATTGCCTTT TCACCTAATA AAGGCGTCAA TGTCAGCGCTG ACTAAGCCAG GATATCCTTT	1920
	ATGAAAGCTA AAATATGCAA AACTAGCGC GACAATACCA TAGACTGCCC ATGCATGAAT	1980
5	CCCCCAATGG AAAAATGAAA ACTGCATTGC ATCATTAAAT GCAGATTGCG TGCCAGCTTT	2040
	ATGAATAGGC GTTAATTTGA AGGCATGACT GATTGGTTCT GCCGTTGTCC AGAACACAAG	2100
	TCCTATTCCC ATACCAGCAC TAAATAACAT AGCAAACCA GACGGCAATG AGAATTCAGG	2160
10	ATCTTCGCCT TCTTCACCTA ATGTAATGTT TGCCTATCTC GAAAAATAAA TATACACACA	2220
	GACAAATAAA ATAATAAAA CGAGCAATAA ATAATACCA GAAAAATGTA GCGCAATAAA	2280
	TGTAGTAATG TTTTGCCTGA GTTTTTCTAA CTGTTTCGGA AATATTGCTC CAAAAGCAAC	2340
15	AAATATCGTA CATATCACTA AAGATACCCA AAACACTAGA CTTACTGATT TATTTTTCAT	2400
	AAATACAAAC CCTTCTGTG TAATGGTAAG TTCATACCCA TAACTGCAAC ATTTTAATCA	2460
20	TTTGTAATTT TATATAGACA CAATTAATAA TGCCTCATCT TTTAAAAATG ATATATAAAA	2520
	CACACTCAAA TTATTTATCA TTGAGCAACA AAGTATTTTA TTGTATTTAA GTAATGCCTT	2580
	TCTAGTGCAT TATTGATTTG ATACCTGCAA AGTTGCCATA TTTCCGTTTA GAATCAATAG	2640
25	TCGCTAGACA CAAAAAATAA GTCGCCTATA CAGTATTTTC TGCATAAGGC GACTTTACTT	2700
	ACTAATCTAT ATATTAATTA CTAATTTTCC AATCATTGAT TGTTTTTCCA ACAATTGATG	2760
	TGCTTGATAT AAGTTTTTCTG GTGATAAACC TTCAAAACT TGTGTCGTTG TTGGTTGGTA	2820
30	ATGCCCTGAT TCTATATTTT TCGTAATATC TTCTAAATAC TCATGTTGTT TAATCATATC	2880
	AGGCGTTCTGA TGAATTGGAC GCGCAAACAT AAATTCATGT GTAAATGTAA TACTTTTTAA	2940
	TTTTAATGCA TTTAAATCTT GATCTTCATT AAAAGCTACG ATAGTCGTAA TATGCCCTAA	3000
35	TGGTTTTATC AGTTCAATCA TAGTATTGTA ATACAAGTCT GTATTATAGG TGCAAAATAT	3060
	ATAATCTACT AATGGAATTT CTTTAAATTG ACGCACTAAA TCCTCTTTAT GATTCAATAC	3120
40	GATATCTGCG CCCATCTTTT CACACCACTC TGTTGTTTCT TGTGCTGATG CTGTTGTAAT	3180
	GACAGTTAAA CCATACCGTT TAGCAATTTG AGTGGCTATA CTGCCTACAC CACCGGCACC	3240
	ATTAATGATT AAGACAGACT TCCCTTCGTT TTCAGCAGGA TTCGTAGAAA TTTTAAATGT	3300
45	ATCAAAAAAC GTTTCATATG CCGTAATACC AGTTAGCGGT AGACTAACCG CTTCAATTAGC	3360
	ACTTATGTTG TGTGGTGCTT TTGCAACTAT AGCTTCTGAC ACCAATTGAT ATGTCGCATT	3420
	TGATCCTTGT CTATTTGGCG ATCCAGCATA AAATACAACG TCACCCGGAC TAAATAATGT	3480
50	AACGTCTGGT CCGATAGCTT CAACAGTACC AATAGCATCA AACCCAAGTA CACGAGGTGC	3540
	TTGAGTGACT TCCATTTGTC GTTGCTTTGT ATCTACAGGA TTTIACACTAA TGCTATTTAC	3600

55

	ATTTCTTCT TCCAATTTAA AGGGCTTCTC AAATCCTATC ATTTTCATAT CGTTTCACCT	3720
	CATTTATGAA CTTATTTCTT ATTATACAAA ATAGAAGCCA TGTGTGCTTA TATCGCAGCA	3780
5	TCATGACTCC TTTTTCATTT GAATATATAA ATAATTACAG ACGACTTTCG TATTAAATTT	3840
	TAGACTTATT TCTACCATGT TGCTGAACAA ATTTACTTTA GATAAAAAAT TATTAAATTT	3900
	TGGTCAATTA ACAAAGTTAG TTTGTTAAAA CGTgATACTT TATTATTCCG TTACTTTAAT	3960
10	AAC TTGTTTA CCAAAGTTAT CGCCAGTaa TAAATTTTAA AATGCATGTG GCGCATTTTC	4020
	AAAACCATCT TCAATGGTTA CTTGTGACTG AATTTTACCT TCTTGAACCC ATGTTGCAAG	4080
15	CTGTTCACTA GCTTCTTTAA AAGCATTAGC GAATTCACCT ACCAAGAAGC CTCTCATCAT	4140
	TACTTGCTTC TTAATAAGCG TACCTTGAAT ACGTGGTCCG ATATCGGCTT CAGGATGATT	4200
	ATATGACGAA ATTGCGCCAC ATACTGGTAC ACGTGCAAAA CGATTTAAAT GCTTAAATAC	4260
20	TTCATCGCCA ACTGTTCCAC CAACATTTTC AAAATAAACA TCAATACCAT CTGGTACTGC	4320
	TTGTGCTAAC GCTTCTGCAA AATCCTCTTT CTTATAATCA ATACCAGCGT CAAAGCCCAG	4380
	TGTCTCTGTT AAATAATTTA CTTTTTTGTC GCCACCCGCA ATACCTACTA CACGGCAACC	4440
25	TTTAATCTTA GCAATTTGAC CTACAACCTGA ACCTACAGCA CCAGATGCAG CTGAAACCAC	4500
	AACAGTATCA CCGGCTTTAG GTTGTCCAAT ATCAAGCAGA CCATGATATG CTGTTTGTCC	4560
	TGGCATTCCCT AAAACACTTA AATATAAATC AAGTGGTACA TCTGTCTGTT GAACTTTAGT	4620
30	AATTTGATCC GCTTGGACAT GATTAATGAT TCGCCAAGGC AACATACCTA CAACGACATC	4680
	TCCTTTTTTA TAATCTGCGA GTGTGCAATC AATTACTTTT GCAACGACAT GGCTAACAAAT	4740
	CGGTTTACCA ATTTCAAAG GCTGTACATA CGAATCTGCC TTAGTCATAC GTCCTCTCAT	4800
35	ATATGGATCC ACTGAAATAT ACAGCGTTTG TACAAGTACA CCATCGCTCT CAAGTTTaGG	4860
	CGTGfCAATC TCTTCaATTT TGAATGTATC CTCTTGAGGC ATGCCKTCTG GTATTTTGT	4920
40	AAAAAGAATT TGTTTATTTT GCATCATTA TCACTTTCT TATTTGAAA CTTTTACTTA	4980
	TTTGTTACTT AAGCGTTAAG TTTGAATTGT GTCTTCGTGA TGTCTGTATG CAAATACATT	5040
	CTTAGTTGTT ATATTTTGAC TTAAGCACTG ATTCATTCAT GTAACCTCAA CCACATTATA	5100
45	TTTGCTATAA TCATAAATTT AAAATGTTAC GACTTAGACA TTTTATGGAA ACTCTCAAAC	5160
	AATAGATAAT TTTTGAAAAG CTCTAATATT ACAAGCTTTT TTGCCCCAGA AAAACTAGCA	5220
	GTTGCTTTAT TTTTCCATA AGAAGTCGAT TAACTCATT GCAACATTTT CATTCTCATG	5280
50	AAGCTGACTA TGTTGTGCAG GCTCACCTTC ATATTTAGAT TCTCGATAAC TTTTCGGACT	5340
	ATTTCCCACT AAATATTTTA ATGATTTGCA AGAACTATTA GAACTCTGC CGTCTGAATG	5400

55

## EP 0 786 519 A2

	ATCTTTTAGC	ACGCGTAATT	GCTGATAAGG	TTGATTCAAT	CGACTTGGTT	TACCATCTTT	5520
	ATCAACTGTA	ATTTCAATGA	CATCTTCATT	CATATTTAAA	ACACCATTAA	ATGTCCCTGC	5580
5	AATATTCACT	TGTTTGTTTA	ACTGTGGCAG	TGACTTGTCT	TTACCATATG	TCATCATATA	5640
	TTGTGCAAAT	GTTAAGTTCC	CCATTGAGTG	ACCGACAAAG	TTGAATTTAT	CGAAATTGTA	5700
	TTCAGATTGT	AACTTAGTCA	GTACATTTTT	AAACCACGCA	GCATTCTTAT	CCAAATAGCC	5760
10	TTGTCTGTTA	TTTTCAAGTT	CAATTTTCAC	AATAGGATTG	ACTGCATCTT	TTCTTAGTTT	5820
	CCCTTTAAAT	GTCACGTGAC	CATCCTTTGA	AACGTAAGCA	GTGATGATAT	CTTTAGTTAC	5880
	CCCTCTTTTT	TCTGCTTGCT	TCACCATAAA	CTTTTCAGAA	TTGGCACTAC	CACCAAATCC	5940
15	ATGTAAGAAC	AATGTTGGAA	TTGGCTTTTT	AACAAATTGC	TGTTGTTGTA	TTTTAAATGT	6000
	TTGTGCCTGT	CGTTGACTAA	ACACCACCAT	AATAATAGAG	CCTATAATAA	TAGCGACCGC	6060
20	TAACAATGTC	GTAATAATTA	CAAAAATTTT	CTTCACACTT	TTAACTCCCA	TTCATGTCTT	6120
	TTATATAAGT	ATAAAGGATG	TGATTAAAAA	TGTCCTTTAG	TTGATTTTGA	ATACATCAAT	6180
	AACTTTTAAG	ATGACTTTGG	AAAGTTGTCC	GTTAACGTTT	GTTAATTGAT	TGCTTCTTTA	6240
25	GCTTTCAATG	GTGTGTCACC	CATTGATTAA	TATATAAATA	TGTATATGCA	TGTTTAATTT	6300
	ATCTCTTTCT	ATAAATAAAG	ACCTACCAGC	ACTCGACTGA	TAGGCCTTTT	AATATCTATA	6360
	ATTATTTAAT	TTCTTTTGTT	TCGGCTAACT	CTTTGTACCA	ATAAGCACTT	TTCTTAGGAT	6420
30	AACGTTCTTG	AGTCTCAAAA	TCGACATAGA	ATAAACCATA	TCGTTTTTCA	TAACCATTTG	6480
	ACCAAGAGAA	CACATCCATT	AATGACCAAA	TAAAGTAACC	TTTAACATTT	GCACCATCTA	6540
	TAATAGCATC	TGCAATAACG	TTCAAATGTT	GTCTTACATA	ATCAATACGT	GCATCATCAT	6600
35	GAACTGTTTT	TTCAGATTCA	ATAAATTCAT	CTTTATATCC	TAAACCATTT	TCAGTGATAT	6660
	AAATCTTATG	aTAGTTAGGA	TAATCTTTAA	CAACACGCAT	GaTTTGATCA	TATAAACCTT	6720
	GAGGATAGAT	CATCCAGTCC	CAGTCTGTGC	GAGGTACGTC	GACATCAAAT	TCACGTTGTC	6780
40	CGACACCTTT	AAGTTGGTAT	TTAGAACCGC	CTTTATCACC	TGTCGCATTA	TGCGTGATTT	6840
	CAGATTCTCC	ATCGTAACCT	CTCATCCAAT	CACTCATGTA	GTAATTGATA	CCTAAGAAGT	6900
45	CGTTTAAATC	TTTGGCTGCA	TCTAAAATGG	CATAATCTTC	ATCTGTAATG	TTTAATTTAC	6960
	CGCCATTAAC	AGATAAGATA	TGTTGCACAC	CTTCCATCGT	TTACGAGAA	TACTTACCTA	7020
	AATATGTTGC	ATCTAAGATG	AATTTATTAT	GGATGATATC	TTCTAATTCT	GCTGCACGAA	7080
50	CATCTTCAGG	ATTTGATGGA	TCGAACGGAT	ATTTTGTTGG	CAATGCGTGT	ACAACACCAA	7140
	TTTCTCCTTT	GTATCCGCCA	TCTTTAAATA	ATTTTACTGC	TCTAGCATGA	GCCACCATCA	7200
55							

## EP 0 786 519 A2

	CTACTAAATA TTGACCATCA CCAATAGGTC CAATTTTCATT GAATGTAGTC CAATATTTTA	7320
	CTTCTGGGAA TTCTTTAAAA CAATATTCAG CATAATCTAC AAAGTAGTCA ATCGTTTTAC	7380
5	GATTTAGAAA ATCGCCATCT TTGTGTAAcA CTTCTGGTGT ATCAAAATGA TGCAATGTTA	7440
	CAAATGGTTC AACATGACGT TTATGACACT CTGCAAATAA CTTATGGTAA TACTCAACAC	7500
	CTTTAGGGTT AACTTCGCCA TATCCATTTG GGAAGATACG AGACCATGCA ATTGAAATTC	7560
10	GGATACCATT AACACCGAAT TTTTCACTTA ATTCTAAATC CACTGGATAT CTGTTATAAA	7620
	AATCACTCGC TGGTTCGCA GTGTACCAAT AGTTTTCTTC TAAATACGTA TCCCATGCTA	7680
	CGCGACCTTT ACCATCCGTA TTGTTCGCAC CTTCTGCTTG ATATGCTGCT GTTGCTCCAC	7740
15	CAAAAATAAA ATCTTCAGGT AATGTTTTAG TCATATGAAA AACTCCTATT CTTAATTTTC	7800
	AAATGTGTGT TGAACGAAAT CAAGGGCTGC TTGGCCATCT CGTGTCAATT TGATATATTC	7860
20	AGCACCTTGA GTCTTCGCTA ATTTAATACC TAATCTATCT GTATCTTGCT TAATATCTTC	7920
	ATAGTTAGAC GCAACTTGTG GCGCTAAAAT GATTAATTGG TACTCTTTCA TAATGTCCAT	7980
	ATGTGCGCCA TATCCGCCAG cTGCCGCTTT CACTGGCACA TGATATTCTT CAGCTGCTTT	8040
25	ATTAAGTGCA TTGGCTAATA ATCCACTTGT ACCACCACCG GCACAAAGTA CTAAGACATT	8100
	TGTTTGTTCT GTGATATTTG AAGCTTTAGC TGCATCGTCT GATACACCAC TTGCCGCTAA	8160
	AATTGAATCA GCTTTTTTCG TATCAAAGTT TGCTGCAACT TTTTCTTTTA AATCTGAATT	8220
30	ACTTTCTTTA CGTCCTTCTT CTTCATCAAG AATTTCACTA TCATAAACTT TTAGGAATGG	8280
	GTAGTAAATA ATAATATCTA CAACAATCAA AGTAATAGCT AGTACGAATG ACCATAAACC	8340
	AAAACCTGTA CCCATGATAA TGCCCAATGG ACCTGGTGTT GTCCAAGGTA AATTCACACT	8400
35	AAACTATTTC ATTCCTAACA CTTCAACGAA AAGTTTGAAA ATCCATACGT TAACAATTGG	8460
	TGCTAATACA AATGGAATAA AGAACACAGG ATTCAATACT AGTGGTGCAC CAAATAAAAT	8520
40	TGGTTCGTTT ACACCAAAGA ATGTTGGTAC AACTGATGCA CGTCCAATCG CTTTGTTTCG	8580
	TTTAGATTTT GTCATCCACA TAAACATGAA CGGGACGACC AATGTTGCAC CCGTACCTCC	8640
	AAATGTAACG ATAAACATTT GTGTACCTGA TGTAATAATT TTATCTGCGT GTTCTCCAGC	8700
45	TTGAAGCAAC TTGAAGTTTC CTTGATATT CGCATATGTA ATGGCTGCAA TTGCTGGCTC	8760
	TACAATTGAC GGACCATGAA TACCTACAAA CCAGAATAAT GCAAAGGCAC CAAAGATAAT	8820
	TGTGACACCA ATCCATCCAT CTGCTGCTGT AAATAATGGT TCGAATAATT TTAAAATACC	8880
50	TTCCGCTACA TTTGATTTAA AGCTGTTGCG AATGACTAAA TCTAATGCAT AAAGAATGAT	8940
	GATTACCGCT GAAAATGGAA TTAAGTCCTT AAATACTTGT GAAATATTCG GCGGTACTTC	9000
55		

## EP 0 786 519 A2

	AAATGCTGAT AAGAATGCTG TTAATAAACC TTTAGTTCCC ATAAATGCAC TTAAGAATCC	9120
	ACCATCTTTG GCTGGATCAG AAGCTAAGAA CAAGAATCCA CACATCGCTG CTAGCATTGT	9180
5	AGAAATAAAG TTAATTTGAT TTGTACTTTC TAGCTTACGG TTAAATGAAT CTGTTAAAGA	9240
	TTTCGCTGTC GTTCCTGCTA CTAAAAATGC TACAAGCCCC ATCGTATAGT TATATGGTTT	9300
	CATTAAAAATG GCTTCCATGC CTTTATCCCA TTTAAAACCA AATATATTTG GCACATATGC	9360
10	AATTAATAGA AAGATACTTG AGAATAAGAT GACAGGCATT GCAGAAATAA ATCCATCAGC	9420
	GATGGCTCTT AAATATATGT TACGTGATAA TTTCTCGAAA AATGGCTTCC CTTTTTCAAT	9480
	TTGTGCGATC AATTTTTGCA TCATTGTCAT CACCCTCTTT TATAAAATTC TAATAAATGC	9540
15	TTCAATAAAT CTTTCAGTAA AATTGTTGTC ATTAAATGAT CTTGACCATG CATCATCGTT	9600
	ACACTGTATG CAATATCATC ACCTTGCCTG TCTTTAGCCA ACAGGCTTGT TTGTGCTCTA	9660
20	TGCGCTTCCG CAATGCAATT GTTTCCTTCT TCAATCAGTG CATCTGCTTT TGCAAAATCT	9720
	CCAGCTTGAG CTGCTGTTAA TGCTTCTAAA AACTTAGAAC GTGCATCCCC TGCAAATGCA	9780
	ACAATTTCAA AACCTAATAA TTGGACTTCT TCTCTATTCA TAGCATTAAAT CCCCTTTTAA	9840
25	ACTTATTTTC TTTGTTTCCA AGATGTCGCA GTATCTTTTA ATACTTTATT TAAGTCATCA	9900
	ATATTTTGA AACCAGTTGT ACGTAACCAT TCACGAGCAG CATCTTCACC TTGTTCAATG	9960
	TATACTTGAA CAGCACCAGA CCATGTAGCA CGGCCACAAA GTACCCCGTT GAATTTAGCA	10020
30	CCAGCTTCGT GAGCAAATTT TAAAGTTTCT TGGATAAATT CCGCAGAAAC ACCAGCACTT	10080
	AAGTAAATGT ATGGTAAATG AGTTGCTGCA TCTTGATCTT TAAAGTGTTG TGCCGCTTCC	10140
	TCTTTTGAT AAACCACTTC ACCTTCAGCA AAGCCTTCTA CATATTTTCA GTTTACTGGT	10200
35	ACTTCAACTT TCAATACATC AACGTAAAG CGTGGTCTG AGAATAATTT CATTGCTTCG	10260
	TTAACTTTTC TAGGCTTAAC TTTTGCGAAT TcAACAcTAC CGTTATcAGG AATGTTGTCA	10320
40	TCGTATGTTA ATACTTCTAA AAAGAATGGA ATATCTTCTG CAACACATTC TGAACCGATT	10380
	CTTTCAATGT ATGCTTTCTT TTGAATGTTA ATTTCTTCAG CATCATCAAC ATCATAGTAA	10440
	AGTAAGAATT TAACAGCATT TGCGCCTTGT TCTTTTAAAC GTTTTGCAGA CCACTCTACT	10500
45	AAACAGTCAG GTAAACGACC TTTAGCGTTT ACGTCATATC CAGTTTTTTC ATAAGCAAGT	10560
	AATAATCCAC AATCTTTGTT ACGTGCATCT GAAGCTGGTA AACCATATTC AGGATCTAAT	10620
	AAAATTGAAG ATGCATATTG TGTTAATTCT TCCGCAACTA ATACTTTTAA TTGTTCAATT	10680
50	TGAGCTACAG TTGGTTCTTC AGTTTGATGT TTTGCCATCA TGCGTTTTAA AGCACCACGT	10740
	TGGTCAAATG CTAATGCAGA AATGATACCT TCGTTGTTAC TTAATTGTTT AATTGATGCG	10800

55

EP 0 786 519 A2

	TCATCATP	TATTTAAATT	GACATAACCT	GTTTGTGCTT	CTTGTGCATT	CAGCATGCCT	10920
	AAAGTATT	CTTTTTTTAG	TAAATCGTGG	TCGTTTTTCAT	GATTAAGAAT	TGCTGAAGTA	10980
5	ATTCCAGC	CTGTAGAATC	ACCTGAACCA	ACCGGATTTA	ATACACTTAT	TGTCGGAATA	11040
	TTCACCTC	AGAATGTATG	ATTGTGCTTA	GCGAATGCAC	CTTGTGCACC	TAAAGACACA	11100
	ATAATCCA	CAATCCCTTC	GAATAAGGGT	TGTGACACTG	CCTGTTTCAA	ACTTTCTAAA	11160
10	CTTTCATC	GTGGCTGGTT	AAGCAATTGA	TATAGTTCAG	AAATGTTTGG	TTTAATGACT	11220
	GTAGGTTT	ATGGATTTTC	CAAACTGTGT	TGCAAAGTtG	CACCCGAGCA	ATCTAATATC	11280
	ACAGGCAC	CTTTGTTTTG	GCATCGTTCA	ATGATTTGTG	CATAATAATC	TTGATTTAAT	11340
15	CCTTTAGC	AGCTACCTGA	AATAGCAACT	GCTTCAACTT	TTTCTAATAA	TTGTTCAAAA	11400
	TGTTTAAT	ATCCTGCAGC	CTCTTGATTA	TCAATCTCCG	GTCCCTGCTC	TAAAATTTCT	11460
20	GTTTGTTC	CTTCATGTAA	AATTGCAATG	CAGTTTCGTG	TTTCACCCTT	AATGTTATAA	11520
	AATGCATG	TGATGTCGGC	ATGATCTAAT	TTTTTAGCAA	TAAATTGACC	TAATTCACCG	11580
	CCAATAAA	CACTCGCAAG	GACTGGCTCA	CCTACTTGCG	CAAGTACTCT	TGTTACATTT	11640
25	AAACCTTT	CACCAGCTGT	TTTACTTACT	TCTTGAACAC	GATTAACATC	ATCTAATTTT	11700
	AATGCTGT	ATGGGTATGA	AATATCAACG	GATGGATTTA	ATGTTAAAGT	TAAAATCATA	11760
	TGTGTCGT	CTTAATCGTG	GTATTGCGCT	CTGTCCCAT	TTTCTAAGAA	TTTCATCAAAG	11820
30	AAATGTGG	CAGCTTGATC	TGCATTGCTT	GTTTCTAAAT	GTTTAATTTT	AGCGATTAAT	11880
	TTTTTGTCT	CTTCAGTTGG	TTTATATTCA	GCATTAATAA	ATGCATCGAT	AATATCGCAC	11940
	ATTAATAACT	CACCTATAAT	ACGTCCACCG	AAGCCAATAA	CGTTCGCATT	TAATTCCTCT	12000
35	TTAGCGTATA	ACGCTGATGT	CATATCACGT	ACTAGTGCTG	AACGAACGCC	AGGTACTTTA	12060
	TTTACAGCAT	TGTTAATACC	AACACCTGTT	CCACAAATAC	AAACACCTAA	GTCTGCATTA	12120
40	CCGCTAACAA	CTTGTTTCGCC	AACTTTTTTA	CCAAAAATTG	GATAATGTGT	TCTTGTGAAA	12180
	TCGTATGTT	CTACGTCAAT	GACTTCATGT	CCTTTTGATT	TTAAAAATTC	AGATACACGC	12240
	ATTTTTGTAT	CTGTAACAAT	ATGGTCGCAT	CCTAATGCAA	TCTTCATAGT	AATTTTTCCT	12300
45	CCTTAGCACA	TTTTATTAAG	CATATCTACG	CGGATTTGGT	GTCTACCACC	ATCGTATTTA	12360
	CCTTCAACAA	AACCTTTAAC	GACATTTTTT	GCTAATGTGT	CTCCAACAAT	TTCAGATCCC	12420
	ATAGTGATCA	TTCTTGAATT	GTTATGGCCT	CTAGTCATAT	ATCCAGAGCG	TTCATCTGAT	12480
50	ACTTCAGCAG	CAATCATGCC	TTTGATTTTT	GTAGCAACCA	TAAAGCTACC	TGCACCAAAT	12540
	GCATCGATAA	CAATACCTAA	GTTACCTTCT	TGACTTTGAA	CATCTTTTGC	TACAGCCAAA	12600

55

TCTAATAAGT ATGATTTGAT GACTTCTTTT AATCGTTTGC CAGCTTCATC TGAACCAATA 12720  
 ATAATCGCCA TAATAAGACT CCTTTTACT TTAATTTTGA AATACCTTTC TTAAAATGTG 12780  
 5 ACATATTTAT TTGTAGGTTA TGAAAATCTT GAGAAAAGGC TTTCAATTTG ATTACGTTTA 12840  
 AATTATAAAC ATAAACAAAC AATAAAATCAA CATAATATGT TTATAATATG TTTGTTTATG 12900  
 ACGTATTTTC AAACAATAAG TGAACATTCA TATTGTGGTG TTGTTTAAAT TAGGTATTCTG 12960  
 10 TCTGAAATTG TAGTAAACT TTGTCGAGGT TCCCGTTGac ATAAATTTGC ATAAAAAaA 13020  
 GCCCATAAAT GAATGCAAAT TCACATTAC TTATGAGCAT ATAGATACAT ATTTTAACAA 13080  
 TGCAATTATA CTTTAAATTT AGTCGACTAC TTCAATATAT GTTTAATCG TTTCTACTTT 13140  
 15 TTCTTCATCT TCATAGTCCA TGACCACTGC AGTCAATTCG TTTAACTGAC AAAATGATGT 13200  
 AAAATCTTCT TTGCCAATT TCGTATGATC GATTAAACAAG TATTTTTCAA TTGAATTACT 13260  
 TAGTGCCAGT TGTTGCGTAT AGGCTTCATC TAATGTAGAT GTCATCACAG CACCTTTATT 13320  
 20 TACTGCGTTA CTAATAAGA ACATCTTGCT AAATCTTAGT TTTTCCAACA TGGCGTTCGC 13380  
 cATTTACCT ACAATGCTT CTGTAATATG GCGCATTTCA CCACCAATTA AATAGACACG 13440  
 25 AAAATGTGCT GTTTGTTTTT CTAACAAAAT TTTATACACC GGCAACAAT TCGTAATAAT 13500  
 TGTGAGCGTA TGATGATTGA CTTCTTCTGC TAATAGTTCC ACTGTTGTTT CTGGTCCGAA 13560  
 AAACAAAGTA TCCCATCTT CAATTAATGA TGCAGCTTTT TTAGCTATAA ATCGTTTTTC 13620  
 30 TGCAATTTGA CGGGTATGTT TTTCTTTATG CGATATTTCT TTATACTGAA ATGTTGAATT 13680  
 ACTGCGTGCA CCACCATGAA TCTTCGTTAA AATCCCTTTA TTTCCAATT CAATTAAATC 13740  
 TCTTCGAAGT GTCATATCAG ACACATTTAA ACCTTCGACG ATTTCAATCG TTCTTATCGT 13800  
 35 GCCCTTTTTA TCACTAGTT TAGCAATTTT GTCCAAACGT TCATGTTTAT TCAATGTAAA 13860  
 ATTGCTC 13868

(2) INFORMATION FOR SEQ ID NO: 174:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4549 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

50 TTAAGTCAAC TTTGTCTATA CGGTTTGGAT CtTCTaCCCA ATGTCTTATA AAAGACAATC 60  
 CCGCACCTGA AACATAACTC ATGAAATAAG AAAATGGTAT ACCATTAATT TGATCATTTT 120



	AATCTTTACC CATAACGAAAC ATCAATTGAT AAAATGCGAT GTCTTTTCT ATCATTCTA	240
	TTAAACGGT CATAATTTGA TGTATGTTAT CCGTGGATAA CTTAACTGCT CCATTAACT	300
5	TCTCATCATG AATGAAGTCT CTTATTTCT CCAACTGCTG GTCCTCTAAT TTTCAAGCA	360
	AATCATACTT ATCATAATAA TGGGTATAAA ATGTACTACG GTTAACATCA GCTAAATCTG	420
	CAATTTGTTG CACAGTAATC TCTTCTAATT GGTGTTGATG TAAAAGTTCA ATAAATGCAT	480
10	TTCTCATGTC AACTTGTGAT TTTCTAATAC GTCGATCTAT AGTCATTTAT ATCAAGTCCT	540
	CCCCAATGAT TATAAACGTT ATGTTTCATTA TCCCACAAAT CTCCAACATT GATGATTGGC	600
	ACACAATGTT TACCTGTTTA ATATAGGTGA TACAAACAAA CAGAAAAAGG TGATAACAAT	660
15	GAACCAACAT TTAGTAGGAA ATCCAAAATT AACTGTAAC TATGTCAATG AAGTTAAAGC	720
	CGGTATTAAC CACATCGTTG TCGACAGTGT TCAATATGGA AATCAAGAAA TGATTATGGA	780
20	AAAAGATGTC ACTGTGGAAA TGGCGGATGG CGAAAAATTA TATATTAATA TTTTCAGACC	840
	AAATAAAGAT GGCAAATTC CTGTAGTTAT GTCTGCAGAT ACTTACGGTA AAGATAATAA	900
	GCCTAAATC ACAAATATGG GTGCCCTTG GCCAACATTA GGTACCATTG CGACATCTAG	960
25	TTTTACACCT GAAGAATCAC CAGACCCAGG ATTTTGGGTG CCAAATGATT ATGTTGTAGT	1020
	TAAAGTTGCA TTACGCGGTA GTGACAAATC CAAAGGCGTC TTATCTCCAT GGTCAAAAAG	1080
	AGAAGCGGAA GATTATTACG ATGTGATTGAA TGGGCAGCAA ATCAGTCATG GAGTAATGGA	1140
30	AATATCGGGA CAAATGGTGT TTCTTATCTT GCGGTGACTC AATGGTGGGT CGCATCTTA	1200
	AATCCACCAC ATTTAAAGC AATGATTCTT TGGGAAGGCT TAAATGATAT GTATAGAGAA	1260
	GTAGCCTTTC ACGGAGGTAT mCCAGATACT GGCTTTTATC GTTTCTGGAC TCAAGGTATT	1320
35	TTTGCGAGAT GGACAGATAA TCCAAATATC GAAGATTGA TTCAAGCACA ACAAGAACAT	1380
	CCTCTGTTTC ATGATTTTTG GAAACAGCGT CAAGTGCCAT TATCACAAT TAAAACACCT	1440
40	CTACTAACAT GTGCTAGTTG GTCTACACAA GGTTCGACA ACCGTGGCTC TTTGAAGGA	1500
	TTTAAACAAG CTGCATCTGA AGAAAAATGG CTATATGTGC ATGGACGTAA AGAGTGGGAA	1560
	AGTTACTACG CTAGAGAAAA TCTCGAACGC CAAAATCAT TCTTTGATTT TTACCTTAA	1620
45	GAAGAAAATA ACGATTGGAA AGATACGCCT CATGTCAATT ATGAAGTTAG AGATCAATTT	1680
	TATAAAGGCG AATTCAAATC AGCGTCACGT GTCCCTTTAC CTAACGCAGA ATATACACCA	1740
	TTGTATTTGA ATGCTGAAAA TCACACATTG AATCATGCAA AGATTAGTAG CGCGCATGTC	1800
50	GCACAATATG ACTCTGAAGA TAAACAACAA GATGTAAGTT TTAAATATAC GTTTGACAAA	1860
	GATACTGAGT TAGTTGAAA CATGAACCTA AACTATGGG TAAGCACTAA AGACTCAGAT	1920
55		

	CCTGATTTTA ATCATATTGA AAATGGTCAA GTAGCTACTG GTTGGTTACG CGTATCACAT	2040
	CGTGAATTAG ATCAAGAAAA ATCCTCAATC GCGCAACCTT GGCATAAACA TGAAACAGAA	2100
5	TTAAAGTTGT CACAAGATGA GATTGTACCT GTTGAAATCG AATTGTTACC TTCAGGCACG	2160
	CTATTTAAAC AAGGCGAAAC ATTGGAAGTT GTTGTAAGG GTAGTGAAAT TGTAATTGGT	2220
	AATAGTACTC CTGGCATGAA AACACGTTAT GAACATGAAG AAACCGTAAA TAAAGGCATG	2280
10	CACATGATTT ATACTGGTGG TAAATATGAT TCACAATTAA TCATTCTCTAT CGTTAATTGA	2340
	TATGCAGCAA TTACGGTCGC TTTTGATTAA AAGTGACATA GTGATAGGAC TGTATAACAA	2400
	GAGAAAGCCA CACGCTTGGA ATCTTAAACC AAGGTGTGGC CCTTTTATT ATTGATGGCT	2460
15	ATTTAATTTT ATAACACTAT CGTATTTTCT TTTTCATGAA TCATTTCAAT AATGACATTA	2520
	TCTTCATTCA TTAAGTCTAC TTTAGGTGCA TGGTTTTTAA TTTCTTCTTC ATTCAACTGT	2580
	GCATAAGTCA TGATTATGAC TACATCGCCT ACTTCAACAA GTCTTGACGC TGCACCGTTT	2640
20	AAACAAATTT TACCACTACC TCTTTCACCA GCTATTACGT ATGTTTCAAA ACGTGCACCA	2700
	TTATTATTAT TCACGATGGC TACTTTTTCA TTTGGCAAGA TGTCTACCGC TTCCAATATA	2760
25	TCTGAATCAA TCGTAATGCT ACCTACATAA TTTAAATTG ACTCAGTCAC TCTTGCTCTA	2820
	TGAATTTTAG CATTCATCAT TGTCTTATC ACTTTATTC GCTCCAATTA TTATATTATC	2880
	TATTAAACGC GCTTTTGAAA ATTTAACAGC TAAcGAGATA AATATGCGTC CAGTTATTTT	2940
30	GTGTTGTTCT ACTAATTGAG GATAACTATA AACAGCAACT TCTTCAATGC GTTCACTTAT	3000
	ATGTGATTCA AGATATTCAG TAACCCTGTC TATAATTACT TTACTTTGAC GTTCACCGTC	3060
	TTGATACAAC GCTTGTGCTA ATAGCAAAC TTTACTTAAA TGTACCGCTT CTTGTGCTTC	3120
35	TTGCTCCGTT AAATAAACAT TTCTTGAAC TTTGCGCAAA CCATCTGCTT CTCGAACGAT	3180
	ATCAATACCA ATAATTTCAA CGGCATGATT GAAGTCTTTT ACCmTTTGCT CGaCAATAGC	3240
	CAATTGCTGG GCATCTTTTT TACCAAAATA AGCATAATCC GGCATAACAA TATTAAATAG	3300
40	CTTATTAACT ACTGTTACCA CCCCATCAAA ATGCCCTGGr CCGtTCGCTC CTTCTAACAC	3360
	ATCAGCTAAT GGGCCTACTT TGACATCAAT ACCTAATTCA CCTGGATACA TATCTTCTAC	3420
45	TGCAGGATGA AAAACAATGT CCGCTCCTAC TTCTGATACT AATTCTAAAT CTTTATCAAT	3480
	TTGTCTCGGA TAAGCATCGA AATCTTCGTT TGGACCAAAT TGTAATGGAT TAACAAATAC	3540
	ACTCACAATT GTAATATCAT TTGTACTAAC TGATTGCGGT ACCATCGTTA AATGTCCATC	3600
50	ATGTAAGGCA CCCATTGTTG GGATAAAACC AATCGTTGTG CCTGAGCGTT TGGCTGCTTT	3660
	AACAATGTGT TGCATCTCTT TTACCGTAGT AATCAGCTTA GTCATTGTTA TTAACCTCAT	3720

55

5 GATCGTATTG TTTTAAACCA TCCACACCAA CACTAAAATC AGCAAATTGC TTCACAAATT 3840  
 TCGCTTTTATG TTCAACACCA TAATTTAACA TATCGTGATA AACCAATACT TGACCATCTG 3900  
 TACCTTTTCC TGCACCAATA CCAATGACTG GAATTGTTAA GTGCTTGCTA ATTTCTTCTG 3960  
 CTAAATCATT TGGAATTGCT TCAAGTACTA ACGCAACTGC ACCAGCTTGT TCTACATTTT 4020  
 TCGCGTCTAA AATAAGTTGc TCCGCTGCTT CTTCGTTGTC ACCTTGTAAT TTATACCCCA 4080  
 10 TAACGCCAAC ACTTTGAGGT GTTAATCCTA AATGTGCAAC AACAGGAATA CCAATTGCCG 4140  
 TTGCTTTTTTc AATAAATGGT GTAATATGCG CTCCTTCTGC TTTAATTGCA TTTGCATTG 4200  
 TCTCCTGATA AAGCTTTAGA GCATGATTTA AGTCTTGTGT CATAGAGATG CCTACTGCAC 4260  
 15 CAATCGGCAT ATCAACAACT ACAAATGTAT TTGGTGCGCC TCTTCTTACT GCACGACCGT 4320  
 GATGAATCAT ATCTGCTAAC GTCACTTGTA CGGTACTTTC ATAACCTAAT ACAGTCATAC 4380  
 CAAGTGAATC CCCAACAGA ATCATATCAA TACCCGCTGC TTCCACTTGT TTAGCACTTG 4440  
 20 GAAAATCATA AGCTGTTACC ATAGAAATTT TAGTTTGCTT TTGTTTCATA TCTATTAATT 4500  
 GACTTACTGT TTTCAATGTT ATTCAACCTC TTTTGCAGT ATnATTAGA 4549  
 25 (2) INFORMATION FOR SEQ ID NO: 175:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8339 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:  
 35 TTATCTTTTG TTGTTTCCTT AGACAAACGA CTAACCACAT TATAATGGAC TAATTTATTA 60  
 ATTFIATTTA ATTCCATTAA GTTATCCGTA AACTAAGTG AAGATGCGGA GTTCACTCTC 120  
 40 GTTGTACTC TCGTTTTTAA TAAAGCACCT CGTAATAATA CAATCATTCT TCTTATTAAT 180  
 GATGCTTGTC TATATACCTG TGTTCTTTCA GCATAACGCA TATAGTTTTC AAGTACACTA 240  
 TTCGTTATTT GTCTTCATC TACTAAAGAC TCTAATGTTT TkGTTTCTAC ATTAAAAGCA 300  
 45 ATTTTTTGTA GACGTTCTAA TTCTTTAGAG TTTTCATCAT CTTTCTCTAC AGTTTTTAAA 360  
 AATGCTAATT TATCATGATA TTCTTTAATC ACGTTACCAT ATTTAAACT TGTTTCGAAA 420  
 GTAGATTTTT GATTTAGATA ATCAATAACT TGTTCTAATA TATAAATTCT AGCAACTTTA 480  
 50 AACGACATAT TGCCAATTAC TGTTTTAGGT GCAGGTTTCG TTAATAATGG CAATAATACT 540  
 TGCGCAACTA CCAAATAAT AATAACCATA CCAGATGCAA TAAATAATAA GTCGTTTCTA 600

55

	ATTGTTCCAT GCACACCACA TAACGTCATA ATTAAAGCGT ATAAACTTCG CTTTGGTGGT	720
	TTCTCAGTCG TTGGATTATC ATCATCATTT TTAGTCATCA TTTTGGAA TGGACTGATG	780
5	GCTAAATAAA AATAAGGATA TAAGACATAA ACCCAAACAA ATCTAAATAG ATAGACAGCT	840
	AAAGCAACAA CAATAGTGAT GCCTATTAAA AAGATTAAAT TGTGCGGTTT TGTTTTGATA	900
	ATTTTAATAA TAACCTCAGG TACTAAAAAT CCTAATATTG AAAAAACAAA GCCATTTAAA	960
10	ACATAACCTA GTATATTCCA TGTATGATTG TAACTCATTT GCAGTTGTGT ACGTACTTGC	1020
	ATAATTCTGT CACGTTTCGAA ACCATGTACA AGTCCTGCAA CTACTGCTGC AATGATTCTT	1080
	GATGCGTGAA ACAATTCAGC AATTAAATAC GTAACAAATG GTGTTAACAA TTGAATAATT	1140
15	GTAAACATAT TAATGTTTTT ATATCCTCGA CGCATCAATG TTAATCGGAA CCTTACTAAT	1200
	GCCATACCTA TAAGTAAACC AACCCTGCG CCACCAATTG ATGCAATTAA AAACAACCTGA	1260
20	ACAGCATCAA CAAGTGAAAA AGCACCTGTA ACTAATACTC CAACAGCTAT TTTAAATGAA	1320
	ATAATACCAG CAGCATCATT CAATAATGAC TCACCTTCAA GAATTGTCAT TGCTCCTTTT	1380
	GGCAAGACCT TTCCTTTAGT GATTGCTTGC ACTGCTACTG CATCAGTAGG ACAAAGAATG	1440
25	GCAGCAATTG CAAATGCTGC TCCAATAGGT AAATCTGGCC AAATCCAATG AATAAATAAA	1500
	CCTACACCTA TCACAGTAGT AATGACTAAT CCTAATGCCA TCATCATCAC TGGCTTAATA	1560
	TATTTCTTA AATGGACTCT AGAAACATTA ACACCTTCTA CAAATAACAA AGGCGCAATC	1620
30	ATTGTTACCA TAAACAATTC AGAATCAAAA TTAAATTGAA CAGGGATTGG GGTAATAAAT	1680
	AGTAACATGC CCAAGAAAAT TTGTATAAAT GCTAGGGGTA CTTTAGGTAT GAAAGTATGG	1740
	ACAAACGAAC TTAGTATCAC AACAGCTATA AATATAAGAA TTGTTTCAA TATTTCCAAA	1800
35	CTTTCACCTC TCTAAAAAGT ATTGTTTAAT TGAAAATTAA GTATCACATC TCGTTGTAAT	1860
	TATACCTTAG AGGATAAATT GAGTTAGCGA CCACAAAAGC ACTTTAATAT AGATATATGT	1920
40	CTACGATTGC AGTACTTAAA TTGCAATTA TTTAATTTTA TTTTATCACT AATTGTTTGT	1980
	ATAAATAAAC AACTTGCTTT CACATAACAA CATTAACCTA TAATACAAAA AATGAGCACC	2040
	TTAAATCGA CTAACCAATT TCaAGTACT CTTTAAATGA TTAATTTTGA AAACAGATTT	2100
45	TCaAAGCAT TGTATGCTT AACAAATTTAG CCAACACTTC AATCGTTTTG ATACCATTTC	2160
	TTACGATGCT CTTCTCGTTT TTCAGCACGT AATTGTAATG CTTCTGTAGA GTTTTGTTC	2220
	TTTGAACCTA ATAATATTGA TGCATGTGTG TGAGCATCAT TTTTCGATA CATATAAGCG	2280
50	CCGTTGCGAT AAGCAGCGCG AGCGACTAAG TGCATGCCGA CTGGTGAAGT TAAATTAATA	2340
	AAAACAAGTG ACAGTAATAA ACGCACACTG AAAAATCCTG TATTCACAAT AAAATAAATT	2400

55

	CTTAAGAAAA CATCTTGGAA TTTCACGATA CCTATTGCAC TAATAAGAGC AATAAAACTA	2520
	CCTAACAAACA ACATCACAGC AGCAATAAGA CTAAAGATTT CTTTGTATTAT TTCCATTAAA	2580
5	CACATGCCCC CCACCAATAA AGCGTGATAT TGAAACAGAA CTTACAAAAG ATATAATGGC	2640
	AATGAGCATG ATTGAATCTA AGAAAGAAAC GGTGCCCATATA AGTACACTTA ACACACCCAC	2700
	AATTGACATT ACGACAGCAC TTGTTGTATC AAATGTAACG ACACGATCTG CTGTTGTAGG	2760
10	TCCCTTGATT AATCTAAATA AACAGATGAT TAATGCAATT CCAAAAATAA TGAGTGAAC	2820
	AATAATCATA ATATGTGTTA TTGTTTGTAT CATCGCGACA CCTCCAATAT TAAGTCTTCA	2880
15	TAATGCTTAA TACTTCTTAA CAACTATCT TTTTCTTTT CTGACACGTC GATACTATGA	2940
	ATAAAAACT TTTTAGAGTC TTGAGAAATT CGTATTACTG TAGACCCTGG AGTTATAATA	3000
	ATTAAATTTG TTAAAAATGT TATTGACCAA TCACTTGTTA GTCTTGTTTC ATATGAAAGT	3060
20	AATCCAGGGT TCATATCTTT TGTTTTAAAA AGAATATAAT TAATCGTGCT AATGCTAGAT	3120
	GTTATTAATT GATATAAATA AACACCTAAA AATTTAATAG CTACCCATAT TTTTCTAACA	3180
	TAAAAATCAT CGCTGAAAAA CCTGTGTAAT ATATAAATGA CAATTAAACC AATTAGATAT	3240
25	CCAGAAAAGA AAGTCGAGAA TTTAAATGA TCTTCATCTT GAAATAATAC CCATAAGAAT	3300
	GCAATGATAA TATTTAAAC TATTTGATTC ATTTAGTCCT CTCCTTTCAA ATGCGGATTT	3360
	ACAAGTTTTT GATATAATTG ATCACTCGTG TTCAACTCAG TTGCATCACT TGTAACATTT	3420
30	AACACAACAG GTGCAGCAAT TCCGATTGCG ATAACCACAA CTAATAAAT ACTTAAATTT	3480
	CTTTTTTCGAT ATAGCGGGAT TTTCTTAAAA TTAACCTCCT CCCCATCTTT ATCTCCAAAA	3540
	TACATATAAA AAAGTATCCT AAATAAACTG TACATTGCAA TTAGACTAGT AATAATCATT	3600
35	AACGCTAGTC CAATATAATT GCCATTTTGC AATGCACCTT GGAAAATAAG TACTTTCCCC	3660
	GGAAAGCCAC TAAATGGAGG CACGCCGCCA ATAGCAAAAA TCATTATAAT AAACGCAACT	3720
40	CCAAATAAAG GTTCTTTTTT AGCTAAGCCA TTCAAATATT GATATTGTCG ATAGCCTGTA	3780
	ATGTAACTA AACTACCAAT AATAAAAAAT AGCAATGTTT TTACAACAAT GTCATTTACC	3840
	AAATAAAATA TTGCACCATT AATACCTGCA AACGTGTTTG TTCCTAAACC TAAATGATA	3900
45	AATCCTATTG AGATTATGAC TTGGTAAGCT GCAATCTTTT TAATATCTTT ATAAGCAATG	3960
	ACACCTATAG CGCCGATGAC CATAGTTATA GCAGCCATAG TTGCTAGCAA TGGATGTATG	4020
	AGATCATTAT GTTGATCAAA TAGTAAAGTG AAGAATCGAA TTAATGCATA GGCCCTACT	4080
50	TTGGTCAITTA ACGCTGCAAA TAATGCTGCA AGCTCAGTAT TTAACACAGC GTAGGCTTTG	4140
	GGTAGCCACA TAAAAAGGAC CAGCGCTGCT TTCGCACTAA ATGCGACTAA GAAGATTAAT	4200
55		

	AAGTTTAAATG TACCTACTGT TTTATAAAGT AAACCTATAC CTAATAAGAA TAGCCATGAA	4320
	CCAATAATAT TCAAGACAAC ATAAATAATT GCAGCACGTA ATTGTTCTAC AGATTGTCCA	4380
5	AGTGTAATGA GTACAAATGA CGCTAGTAAC ATAATTTCAA ACATGACGTA TAAATTAAAT	4440
	AAATCTGATG TTAGAAAAGA GCCTATCACG CCAACACTTA AAAATAATAT GAACGATGGC	4500
	AAGTGATAAC GATTTGCTTT ATGTTTCGCCA CGCCCAAATC CGTATGCCAT AATTAAAGTA	4560
10	ATCACAAACG AAGCGGTTGT AACCATAATT AAACCTAAAG AATCTCCTAA AAACCTGTATA	4620
	CCAAAGGGCG CTGACCATCC TCCAAAGTCT AGCGTAATTG GACGGTGACG CTGAACATAA	4680
15	ATTAATAGCA TTAATGAAAT AATTGTGGTG ATAGTCATTG TACCTAAGTA TAAATATTTA	4740
	GAAATACGAT CATTATTTTT TAAAAATACA AGGATTAAGG CACAAAGGAA TGGTAATAAC	4800
	ATTGGTAAAA TCAATAAGTT ACTTAGCATC ATCTTCCCCC CTTAGGCCTT CAATTTTCATC	4860
20	TTCTTTTGTT ACTTTATAAG TTCTATAAAC AAGTACAAGT AAAACGCGAG TCATCCCCAA	4920
	CCCTATAACT ATTGCAGTTA GTACAATAGC TTGTAACAAG GGATCAACAA ACAATTGGTT	4980
	TCCACCAGTT ATTAGTGGTT CTGATCTACT AGAACCATAC GTTCCCATAC TCATAATAAT	5040
25	GAGATTACCA GCATGAGTAT ATATTGAAAT TCCGATTACA ATACGAATTA AATTGATTGA	5100
	TAAAATCATA TATGTTCTTA TAAACACTAA AAATCCTATA ACTAGTAATA ATATTAAATT	5160
	CATGATCGAC CTCCGCTAAG CGACAACATC ACTGTGACAA TAACACCAAC AACTGAGAAT	5220
30	AAAATACCTA ATTCAAAAAG TGTATTGTGA CTTACATGAA TTTGTCCTAA AATTGGAAGT	5280
	ATCCAAGTTG TTTTCATATTG AGACAAAAAT GGTTTTCCAA AAAACATAGG TATTATCGCA	5340
	GTAATAGATG ATACCAATGC TCCAATAATC ATTAATAATC TAAAATCAAT CGGTAAACTT	5400
35	TCTAAAACCT CTTCAACATT AAAAGCCAGA AACATTAAAA TAAACGCTGA ACTAAATATT	5460
	AAACACCAA TAAACCCACC ACCAGGATTA TTATGACCTG CGAAGAAGAC ATAGAATCCG	5520
40	AAAGTCAATA AAATAAATAC AACAAGTTTC GTGACCGTTC TTAACACGAC ATCATTCTCT	5580
	TTTCATCTGT CCCCTCCGAT CTTGATAATT TAATAATGtg TAAATACCTA GCCCAGTAAT	5640
	AATTAACACT AATCCTTCAA ATAATGTATC TAATGCTCTA AAGTCACCAA GTATCGCATT	5700
45	TACAATATTT TTACCACCTG TTAGTTTGTC AGCTTTTAAA TAAAAGTCTG ATATTGATGA	5760
	TAAACCATCT GTTTGTTGTG TAATAAAAAT TAATGATACA ACAATAAGTG CCATCAAGAG	5820
	TGATACAGAA ATTTTAATTA TTTCTCTTTT TTTGTTAGCG TTAGATCTTG GCACGTTTGG	5880
50	TAATCTTGAA AAACCTGACAA TAAATAGTAT CGTCGTTATT GTTCAACTA CTAGCTGAGT	5940
	CAATGCTAGA TCAGGGGCTT TCATTGCTAT AAAGAATAAG GTCACAACAA ATCCGATGAC	6000
55		

	GACAGTTACG ATTGCTAATA TAATTTCTAA TGCCCCAAAT TCAGAAACAT GTAACGTATG	6120
	TACTTTAGGA AGTCCaATTC GAATATAACC ATATCCAATG ATAATCATAA ATATGCCTAA	6180
5	GGTCATAATA ATGTACTGGT TTAAACGATC TTGCATAACA CGTTTAAATC GCTTCGTAGC	6240
	AAACTTTTCA AAATGTCGAT ATACCATCTC ATAGCTTTTT GAAACTGAAA TCTGTCTAAT	6300
	TTTACCTGTG AACACTTTTT TCCAATCTAC TTTGATTGCT AGTACACTAC CCAATAAAAT	6360
10	AATGATGATG GTTAAAAGAA GCGGTATGTT AAATCCATGC CATTGCGAAA CATGTGGTGC	6420
	CAATTGATCA ATTTGATGAT TACCACCTGA TACAGCTCTT AATGChAGAA CGATAATCCC	6480
	CTTCCCAAAT ATATnTGGTA CAAAAAGAT TACAGGTACT AGCACCATTA aTATAAGAGA	6540
15	TGGTAAACTA aACAACCATG GTTCGTGGAT ATTTTTTTTA GTAAAAACCT TAGAATCATA	6600
	TTTTGtCCAA AATACTTCTT TTACCATGTA TAGTGCATAT GTGAATGTAA AAACACTCGC	6660
20	AATAACACCA ACAACACGA TAGCTATCAT TGAAATCAAA CTAAATTGGG ATAATTGTCC	6720
	AGTTTGTGTT AATGCATCTA AAAACATTTC TTTACTTAAA AATCCATTTA AAAATGGTAC	6780
	TCCAGCCATA GATAGAGCCG CTATCGTCAT GACTAGATTC ATTTTAGGAA ATAGTTGACG	6840
25	CATTCCACTT AAAATTCGTA TATCCCTTGA ACCTGCTTCA TGATCTAAAA TACCTACTCC	6900
	CATGAAAAGC GCACATTTAA AGATGGCATG ATTCAATTAGa TGAAATAGcG CACCArATAA	6960
	TACmAATACA TAAATaGATG CTATTGCGTC TTGTTGGTGT TGAGCATATC CGCCACCTAT	7020
30	ACCCACCATA GCCATAATCA TCCCAAGTTG ACTGATTGTA GAGTACGCTA GGATACCTTT	7080
	TAAATCCCAT TGTTTTAAAG CTGTAATTGA ACCAAATAAC ATTGTTATTA AACCAACAAA	7140
	CGTAACGATA TATACGTACA TATTGCTAnG ACCTAATAAT GGTGTAAATC GAAGTAATAG	7200
35	AAhGATACCA GCTTTTACCA TCGTGGCTGA ATGTAAATAA GCACTTACAG GTGTAGGTGC	7260
	AGCCATTGCT CTAGGTAGCC AGTATGAAAT GGArattGTG CTGATTTTGT AAATGCACCT	7320
	AATAAAAAACA TAAAAATCAT AGGGATAAAC AATCCATGAT TCTTAATATG ATCTGCTTGT	7380
40	CCTAATATCT CTGTGATGTT ATTCGTTTCCT GTCATGATAT ACAGCATAAT AAAACCAACT	7440
	AATAACGCCA ATCCACCAAA TACTGTAATC ATAAATGATT GAATCGCACC AAATTGACTG	7500
45	TCACCATTGT TATACCAATA TGAnATCAAT AAAAATGATG ATmCACTCGT TAATTCCCAA	7560
	AAaATGTACA TCmATATCGT ATTGTCTGAT AATACaaTAC CAATCATACT GAACATAAAT	7620
	AACGTTAAAT AAAAATAAAA CCTTGGTAAA TTGTCTTTTC GAGAGGATAA ATATTGAGTT	7680
50	GCATAGAAGA ATACTGCAAT TCCAATAAGT GAAATAATAA GAGAAAACAT TAAACTTAAA	7740
	CCATCTAAAC GTAAATCTAA ATTAATATCT AATGTCTTAA TCCATGGAAT AGAGGTAGAA	7800

55

GGTGCAACCA ACGCTATGTA CCCGGCATAT TTAGCCAATG CTCTACGTTT AGACATTAGA 7920  
 AGTATCATCG CCATAATCAC AAGTATAGCA ATTAATAAAT AAACCAAACT CATTATTAGC 7980  
 5 CTCCTTTGTT TCTATAATTG TAATGAAATA TAAATACTAT GTTCACACTC ATTTTCTAAA 8040  
 CCGATAAAAT TTAGTGTTTC AATAGCAGAT TGATGCCCTA AATACTTTTG AATGACTGGT 8100  
 ATAAGTATAC CTTTTTGATA AGCATGATAT GCAAAATGTCT TACGCAATGT CGTTAGTCCT 8160  
 10 ACATTATCTA TACCAGCTTC AATTGATGCT TGGTGAATTA TTCGATATGC TTGCTGTCTA 8220  
 GATAATACTT GATTTGTTTC TAGTGATTGA AAAAGAACGT CTTCAATCGA AAGACTCCTG 8280  
 15 TCCTCTATAT ATTGAAGTAG TTCTTTTCGAT AATGTTTCTG GTAACCTAAT TTTAATCAA 8339

(2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 588 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

CCCgATTTTT TTACGTAATC TAATACATAC GGCAAAATCA ACTTTAATCA AAAAAGACTC 60  
 ATACACAATG CCTTTAAAGC ACATGTATGA GTCCTTTTTA GTAGTTTATA TCAAAAATA 120  
 30 GTTTAATGTA TAAATTAGTT TTTGTTTACA GATGCGTCGT AGATTGATTC TACAGCATCA 180  
 CCTAAAGCTT TATCGAATTC TTCTTTAGAT TGATCAGCTC TTAAATCACT AGCTAATGCA 240  
 CGTGAGAAAC TTGCGATAAG TTCAGCGTTA TCTTTAAGTA ATTCATTTGC TTTTCTCTG 300  
 35 CTGTAACCAC CTGATAATAC AACGACACGA ACAACATTAG GATGTTTCAGC TAACTCTTTG 360  
 TATTAAGTTTG GTTCAGTAGG AATTGTTAAT TTCAACATTA CTAATTGATC AGCATTTAAG 420  
 40 CTATCTAAAC CTTTTTTAAG TTCAGCTTTT AATACTTTTT CAATTTTCAGC TTTGTCTTTT 480  
 GCATTAATAT TAACTTCTGG TTCGATAATT GGAACATAAC CTTTAGCAAT AATTGTTTAA 540  
 GCAACTTCAA ATTGTTGTTT AACAACGTCT TTGATACCTT GTCATTTT 588

45 (2) INFORMATION FOR SEQ ID NO: 177:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2841 base pairs  
 (B) TYPE: nucleic acid  
 50 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55



EP 0 786 519 A2

	ATAGAGTnCT	GGnACTTACT	ATGACATATG	GCGCTAGAAT	GGCTGAGCCA	GGTGAATTTA	60
	CAAAACGTGC	CTTTTAAAT	GGTCGTATTG	ATTTATCTCA	AGCTGAAGCA	GTTATGGACT	120
5	TTATTGCTC	GAAGACAGAT	AGAGCTTCTA	AAGTTGCGAT	GAATCAAATT	GAAGGTCGTC	180
	TAAGTGACTT	AATCAAAAAA	CAACGTCAAT	CTATATTAGA	GATACTCGCT	CAAGTGGAAG	240
	TGAATATTGA	TTATCCTGAA	TACGATGATG	TTGAAGATGC	GACTIONGAA	TTCTTTTAG	300
10	AGCAGTCTAA	AGAAATCAAA	CAGGAAATTA	ATCGTTTATT	AGATAACCGT	GCGCAGGGTA	360
	AAATTATGCG	TGAAGGTTTA	TCTACAGTTA	TTGTTGGTAA	ACCAAACGTA	GGTAAATCAT	420
	CGATGTTAAA	TAATTTAATA	CAAGATAATA	AAGCGATTGT	AACTGAGGTA	GCAGGTACTA	480
15	CTAGAGATGT	CTTAGAAGAG	TACGTCAATG	TTCGTGGCGT	GCCATTAAGA	TTAGTTGATA	540
	CTGCTGGTAT	ACGTGAGACA	GAAGATATAG	TAGAGAAGAT	TGGTGTGAA	CGCTCTAGAA	600
	AGGCTCTTAG	CCAAGCAGAC	TTAATTTTAT	TTGTATTAAA	CAATAACGAA	GCATTGACWC	660
20	AAGAAGATTA	CACATTATAT	GAAGTGGTTA	AAAATGAAGA	TGTAATCGTA	ATTGTTAATA	720
	AAATGGATTT	AGAGCAAAAC	ATAGATATTA	ATGAAGTTAA	AGATATGATA	GGTGATACGC	780
25	CATTAATTCA	AACTTCAATG	TTAAAACAAG	AAGGTATTGA	TGaATTAGAA	ATACAAATTC	840
	gAGATTTGTT	CTTTGGTGGA	GAAGTACAAA	ATCAAGATAT	GACTTATGTT	TCTAATTCAA	900
	GACATATTTT	ATTATTAAAA	CAAGCAAGAC	AAACGATACA	AGATGCGATT	GATGCAGCAG	960
30	AATCTGGTGT	GCCTATGGaT	ATGGTACAAA	TTGATTTAAC	TAGAACTTGG	GAAATATTAG	1020
	GAGAAATTAT	TGGTGAGACT	GCAAgTGATG	AACTCATCGA	TCAGTTATTC	AGTCAATTCT	1080
	GCTTAGGTAA	ATAGTAATTG	AAATAGACGG	AATACCGTCT	TAAGAAGGCT	AGTAAGATAT	1140
35	CAAATAAGGA	GGTTTATATT	GTGGTTCAAG	AATATGATGT	AATCGTTATA	GGTGCGGGAC	1200
	ATGCAGGTGT	AGAAGCAGGT	TTAGCATCTG	CAAGACGTGG	TGCTAAAACA	TTAATGCTAA	1260
	CAATAAATTT	AGATAATATT	GCATTTATGC	CATGTAACCC	ATCTGTAGGT	GGACCAGCTA	1320
40	AAGGTATCGT	TGTTTCGTGAA	ATTGATGCTT	TAGGTGGACA	AATGGCAAAA	ACAATCGATA	1380
	AAACACACAT	TCAAATGAGA	ATGTTAAATA	CAGGTAAAGG	ACCTGCTGTA	AGAGCACTAA	1440
	GAGCGCAAgc	AGaTAAAGTA	CTTTATCAAC	AAGAAATGAA	ACGCGTGATT	GAAGATGAAG	1500
45	AAAATTTGCA	TATAATGCAA	GGTATGGTAG	ACGAACTTAT	TATAGAAGAT	AATGAAGTTA	1560
	AAGGTGTACG	TACAAATATT	GGTACAGAGT	ATTTATCTAA	AGCAGTAATT	ATTACAACGG	1620
	GAACATTTTT	ACGTGGTGAA	ATCATTTTAG	GTAATATGAA	GTATTCAAGT	GGACCAAATC	1680
50	ACCAATTACC	ATCAATCACA	TTATCAGACA	ATTTAAGAGA	ACTTGGTTTT	GATATTGTTC	1740

55

EP 0 786 519 A2

	AAATACAACC AGGTGACGAT GTAGGTCGTG CATTGAGCTT TGAAACAACA GAATATATAT	1860
	TAGATCAATT GCCATGTTGG CTAACGTATA CTAATGCTGA AACACACAAA GTTATCGATG	1920
5	ATAATTTACA TCTATCTGCA ATGTATTGAG GGATGATTAA AGGAACCGGG CCACGTTATT	1980
	GCCCTTCAAT TGAAGATAAA TTTGTTGAT TTAATGATAA GCCGCGACAT CAACTTTTCT	2040
	TAGAGCCTGA AGGTCGTAAT ACAAATGAAG TATATGTGCA AGGATTGTCT ACAAGTCTTC	2100
10	CTGAACATGT GCACGTCAAA TGTTAGAGAC GATACCAAGT CTTGAAAAAG CAGATATGAT	2160
	GCGTGCCGGC TACGCAATTG AATATGATGC GATTGTGCCA ACGCAGTTAT GGCCTACACT	2220
	TGAAACGAAA ATGATTAAAA ACTTATATAC TGCAGGTCAA ATTAATGGTA CATCTGGTTA	2280
15	TGAAGAAGCA GCAGGACAAG GATTGATGGC AGGTATTAAAC GCTGCAGGTA AAGTGTAAAA	2340
	CACAGGCGAA AAGATATTAA GTCGTTGAGA TGCATATATT GGTGTCTTAA TCGATGATCT	2400
	TGTAACATAA GGTACTAATG AACCTTATCG TTTACTAACA TCACGTGCAG AATATCGTTT	2460
20	GTTACTACGT CATGATAATG CTGATTTGAG ATTGACGGAT ATGGGATATG AACTTGGTAT	2520
	GATTTCGAA GAAAGATATG CACGTTTAA TGAAAAACGT CAGCAAATTG ATGCGGAAAT	2580
25	TAAGCGTTTA TCAGATATTC GTATTAAACC AAACGAACAT ACGCAAGCGA TTATTGAACA	2640
	ACATGGTGGT TCTCGCTTAA AAGATGGTAT TTTAGCTATC GATTTATTAC GCAGACCTGA	2700
	AATGACTTAC GATATAATTT TAGAAGTTT AGAAGAAGAA CATCAATTGA ATGCAGATGT	2760
30	TGAAGAACAA GTAGAAATAC AAACAAAATA TGAAGGTTAT ATCAATAAAT CACTACAACA	2820
	AGTTGAGAAA GTTAAGCGTA T	2841

(2) INFORMATION FOR SEQ ID NO: 178:

35	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 3025 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

	ATCTAATTTT AAACCCGGTG ATAAATTGCC AAGCGTGACG CAATTAAAAG AACGTTATCA	60
45	AGTAAGTAAG AGTACTATCA TTAAAGCATT AGGCTTATTG GAACAAGATG GTTTGATCTA	120
	TCAAGCACAA GGCAGTGGTA TTTATGTGAG AAATATTGCT GATGCCAATC GTATCAACGT	180
	CTTTAAGACT AATGGTTTCT CTAAAAGTTT AGGTGAACAC CGAATGACAA GTAAGGTACT	240
50	TGTTTTTAAG GAGATTGCAA CGCCACCTAA ATCTGTACAA GATGAGCTCC AATTAAATGC	300

55

	CGAATATTCT TATTATCATA AAGAAATCGT GAAATATTTA AATGATGATA TTGCTAAGGG	420
	CTCTATCTTC GACTATTTAG AATCAAACAT GAACTTCGT ATTGGTTTTT CAGATATTTT	480
5	CTTTAATGTA GATCAACTCA CTTCAAGTGA AGCTTCATTA CTACAATTGT CTACAGGTGA	540
	ACCATGTTTA CGTTACCACC AGACTTTTTA TACAATGACT GGCAAACCCT TTGATTCATC	600
	TGACATCGTA TTTCAATTATC GTCATGCACA GTTTTATATT CCTAGTAAAA AGTAATAAAT	660
10	ACATAAAAAC GTCTATATCC CAGTTATAAA CTGGAGTATA GACGTTTTTT TACGATAATA	720
	ACAAATGGCTC AAATTGCTAT TATCTTGCTT AGGTTTTTCG TTTTAGAAGA ATATTGCTAC	780
	AAAGACAGGC ACAACTGCTA CAACAACTAC ACCAACTAAC ACTAAAGCTA TACTTGCCAT	840
15	TGATTCTTCT ACAGGTCCTA ATTCTTTGGC TGGTGCTACA CCTAATGTGT GACCACTTGT	900
	TCCAAGTGCT AATCCTCGGG CAATAGGGTT AGTAATTCGG AAAAGCTTTA AGAATTTATT	960
	ACCTAGGGCA TAAATAATGA CACCATTAA AATAACTGCT AATGATGTTA ATTCCTTTAT	1020
20	ACCACCGATA CCAGCTGATA CTGGTAACGC AATCGCTGTA GTTGCTGCTT GAGGTAACAT	1080
	TGATAAAATA ACATCATGG CAAATTGTGC TAACCTCGCA AAAGTTAAAA TAATTAATAA	1140
25	CGCTACAACT GTACCGATAC CAATACCTCC GATGATACGA TGCCAATGTT TAACAAGCAC	1200
	TTACGCTTTT TTATATAACG GAATCGCAA ACAGATTGTT GCCGGTTCTA AGAAGAAGTA	1260
	AATAATGTCT CCACCTATTT TGTAAGTCTT ATACGGAATG CCTGTTAAAT AGAGGAAGGC	1320
30	CACACCAAAT ACCATACTGA CAAATAGCGG TGCGAATAAG AAGAAACGAT TAGTTTTTTC	1380
	AAATAATATG GTCGCTAAGA AAAATGGTAT AACGGATAAC AGTATTCCGA AGTAAGGTGT	1440
	GTTTAgTGCT AAGTGGTTAA TCaTGAGCTT GTGCCTCCTC TATTTTGATC TTTTTTGTGA	1500
35	CTTTGTCACC TTTAGATCTC GAAGTAACTT TCATAATAAT TTgTGTGACA TAGCCAGTAC	1560
	AAATAgGTAA TAGTATTGTT GAGACGATTA TTAGTCCAAT GATTAAAAAT GGTGCTTGGC	1620
	TAATgACACC TAAAGAGTTA ACAACTGAGA TACCGGCTGG TACGAAGAGT AAGCCAATGT	1680
40	TATTTGTTAG TGTCGTTCTT ACTTTTTCGA CTTGCGCTAA CTTAACAGCA CCAGTACATA	1740
	ATAATACAAA TAATAATACT AAACCGATTA CTGATGCAGG CATAGGAATT GGCATAAATG	1800
	ATTCAATTAT TTTGATACA AAGAGTACTA AAGCAATTAC AATGACTTGG TGAAAAAGT	1860
45	GTGCTGGTTT TGATGCGTCT TTTTGTGTGTT TCACGACCAT TGCCTCCTAC GTTTGATTTA	1920
	ACTAAAGTAT AGATGGCTCA CTTCGATTG CGTGATTTTT AGTCCGAAAT ACAAATATC	1980
50	ATAGGTAAAA TGCATAAAAA AAAGGATTAC TGTTAAAGTA ATCCTATCGA CGCTTTAAAA	2040
	TCTTTCATAA ATGAACGTCC AACTTGCATC TTGACACCAT TTGTCAATAT TACCATATAA	2100

55

	TGAATACGTA TAAAATAAGT GGGATTCAAT CGTTTTTCAT AACGATTCAA TGGCTCTGTT	2220
	GTTTCGTATT TATGATTTCGT TGTATGTATG GTTGTAAATAC CATTATGTGT GCCAATCCCA	2280
5	ATAATATTTT GTTGCTTTAA CATGTGAATT TTATCGTCAA TTTCAACAGG TAAGCTTTGA	2340
	TCAAATTCG CCGACATATC ATTCGCAATT GCACTTGCCT TATTATCATC TTTGGCTTTA	2400
	GTCGCACGCA CTTTATTGAC TGCTTGTTCA ATACGTTTTT GACCAAACGG TTTCAAATA	2460
10	TAGTCTGTCG CATTTAATTC AAATGCCTGT ACTGCGTATT GGTCATGTGC AGTTGCAAAA	2520
	ATAATCGCAG GTGGCTCTTT CATCTTTTGA ATCTTAGCTC CTAATTCGAT CCCATTTTCA	2580
15	TCCATTAAAT TGACATCTAA AAATATAATG TCATATTGAT TGATCAGTAG TGCTTCCAAT	2640
	GTTTCTTTTA CATTCTTCTGC CTCATTAATT TCTTCAAAAC CACCAATTTT ATTTAATAAA	2700
	TATGTTAATT CATTACGTGC TAATGGCTCA TCATCTATGA TTAATGCTTT CATATTTATT	2760
20	CCTCCTCTTG TCTTTCATAA GGAAGTACAC ACCAAAAAGT GGTACCGCTC GATGTCGATT	2820
	CAAATTGTAA TGCTGCGGAT TTTCCAAATA ATCCTTTTAG GCGTAAGTTT AAATTTTCTA	2880
	AAGCACTACC AGTTCAGAC TCTGATTCTA CAGATGTnTC TCCCaACAAA TGCATTTTAT	2940
25	CTTTAGAAAT ACCCTGACCA TTATCTTGTA CAATAATACG TACATGTGTT GCAGTTTCTT	3000
	TAATCACTGA CACGTCAATA TCGTT	3025

## (2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1689 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

	ACAGAATTTT ACAGCATTTT TAGATGAAAA AATAAGCCAG TCATAGCGTT GATTTAACAA	60
40	ATGAATATCA AAATTTAGTG GCTTTATATC AATAAAGGGT TTGTGAATAA TTGATACTAA	120
	ATCACTTTGC ATGTCATTTG TTTGTGTCAT AACTACAACCT GGCTTCATAT TTAAACGTCA	180
	CTCCATTATT TAATGTTGTT CATTAAAGCG TTTTATAATT TCATAAGCAC CTTGCTCTTT	240
45	TAATTTGTTA CTCACTGTTT TGCCTAACTC AACCGGATCT GTTCCGTTCA TTGTATATTC	300
	AAATCGTTCT TTACCATCTG GGGTCATAAT TAAACCTGTA AATTCGATTT CGTTTTGATC	360
50	TGAGATTGTA GCATATCCTG CAATTGGCAC CTGACAACTA CCATCCATTT CTGCTAAAAA	420
	CGTTCGTTCA GCAGTCACAC ATTTTGCAAC CTCATCATTA TGTACTTTGC TTAATAATGT	480

55

5 TAACAATGTA TCTCTATCAA GATAAGATGT TnCAATATCA TCTGACCAGC CCATTCTTCT 600  
 TAAACCAGCT GCAGCTAAAA TAATCGCATC ATAATCTTCA GTTTGTAAC TTTCTAATCG 660  
 TGTATCTATA TTACCTCTAA TCCATTTAAT CTCTAAATTA GGATACTTAG ATAATATTTG 720  
 TGCACCACGA CGTAATGAAC TAGTACCAAT AATACTGCCT TCTGGCAATT GGGATAGTGG 780  
 10 TGTATGTGTT TTAGAAATAT ACGCATCAAA AGGTAATTCT CTATCAGGGA TACAACCTAA 840  
 TGTAAACCT TCCGGAATTA CACTTGGTAC GTCTTTAAGC GAGTGTATTG CCATATCGAT 900  
 ATTTTTCAT AAAAGTTCAT GTTGTATTTT TTTAAACAAT AAGCCTTTGC CTCCGACTTT 960  
 15 AGACAATTGT TTATCTACTA TACGATCGCC TTTCGTGACA ATTTCTTTAA TTTCAATTTT 1020  
 TAGATTTGGC TCGACAGCTT TTAATTTATC AATAAATTGC TGGCTTTGTG TTAAAGCTAA 1080  
 TTTACyTCTT CTGGAGCCAA CGACTrATTT ACGCATGTTC AATTCCTCCT AGGAACGGAT 1140  
 20 TGCTCTAGAT TATTTTCTCA ATTCACAAAA TGTGTTGCAA AAAATAAATT AATCATATTT 1200  
 AAGCAAAATA AAATAATGTT ATAGTATATT AAATATCTTG AATTCAACCA TTTGTTGATT 1260  
 CTAAGTAAAA TATAACTTCC ATATAATACT GTAATAATTG AAGAGAGTAT TACCTTCGGG 1320  
 25 TCAATGAATA TACGTTCAAC AACTGAAATT ACACCCCACT GTGTACCTAA AATAATACTA 1380  
 AATATGAGAA TTATCCACCC ACTTAACGTT GAGTAAAACA CAATTGATTC AAGTGTAGCA 1440  
 ACGCTACCAA TTCTAAAGTA TTTTGTATCA AAACGTTTTT CCTTCAAATT ACGGTATTGC 1500  
 30 ATGATATACA GTAATGCATT GACAAAAGCT AAGGCAAAGA AGACATAACT TAACACAGCT 1560  
 AGACCGATAT GGAATAACAG TAACTCGTCT ACAACAGCAA TTTTCTGAAC CTTATTAGTA 1620  
 TAATGTGTCG GTTGAAATGT ATTCATCCCT AAnAGTGTTA ACCCTATTAA ATTCCAAGGA 1680  
 35 AAAACACAG 1689

## (2) INFORMATION FOR SEQ ID NO: 180:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1209 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

nTGGnTGGCT TTTCTATTG GACCAAATGG ACCnTTTACC TGGCCnTTCC CAGGACACCC 60  
 50 CGCTTGTGCC CACATTCCAA TCGGAAAAGG TGTATGTGGT ACAGCCGTTT CAGAACGTCG 120  
 TACACAAATT GTAGCTGATG TTCATCAATT CGAAGGACAT ATCGCTTGTG ATGCTAATAG 180

55

CGATGCCCT ATAACGGATC GATTGTGATGA CAATGACAAa GAaCATCTTG AaGCAATTGT 300  
 TAAAATTATT GAAAaGCAAC TCGCATAAAA GGACATCAGC ATTTTCAATA AAGTGTGAC 360  
 5 AGTTAGCAGG AAAATGTTAC AATAATCTTT GTGTGAATTA ACGAAAGTAG CAGTTGTATA 420  
 TTATTGAGCG CTATGTTGTT CCCAATGCGG ACGTGTACG TAACTGTCGC TATAAGGTGA 480  
 AGACACATAA AACAATATAT CTTAGTAAGC ATGCAACACT CTTTTTTGTT TATTCATAAC 540  
 10 AACAAAAAAG AATTAAAGGA GGAGTCTTAT TATGGCTCGA TTCAGAGGTT CAAACTGGAA 600  
 AAAATCTCGT CGTTTAGGTA TCTCTTTAAG CGGTACTGGT AAAGAATTAG AAAAACGTCC 660  
 TTACGCACCA GGACAACATG GTCCAAACCA ACGTAAAAAA TTATCAGAAT ATGGTTTACA 720  
 15 ATTACGTGAA AAACAAAAAT TACGTTACTT ATATGGAATG ACTGAAAGAC AATTCCGTAA 780  
 CACATTTGAC ATCGCTGGTA AAAAATTCGG TGTACACGGT GAAAACCTCA TGATCTTATT 840  
 20 AGCAAGTCGT TTAGACGCTG TTGTTTATTC ATTAGGTTTA GCTCGTACTC GTCGTCAAGC 900  
 ACGTCAATTA GTTAACCACG GTCATATCTT AGTAGATGGT AAACGTGTTG ATATTCCATC 960  
 TTATTCTGTT AAACCTGGTC AAACAATTTT AGTTCGTGAA AAATCTCAA AATTAAACAT 1020  
 25 CATCGTTGAA TCAGTTGAAA TCAACAATTT CGTACCTGAG TACTTAAACT TTGATGCTGA 1080  
 CAGCTTAACT GGTACTTTTCG TACGTTTACC AGAACGTAGC GAATTACCTG CTGAAATTAA 1140  
 CGAACAATTA ATCCGTTGAG TACTACTCAA GATAATACGG TCAATACCAA CACCCACAAT 1200  
 30 TGTGGGTGT 1209

## (2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 698 base pairs  
 35 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

AAATCCCTTt GTtaAaGtsC AAAtTTTTCc AACrgCTTTA AtArGACCCA TATTACctTC 60  
 TTGGATTAAA tCmAGGaATG AcATACCACG ACCaCGTATC TTTTAGCAAT ACTTACAAC 120  
 45 AAACGTAAGT TCGCTTCTGC AAGTCTTGAT TTTGCTACTT CATCACCTTG TTCAATACGT 180  
 TTGGCTAATT CGATTTCTTC TTGTGCACCTT AATAAGTTAA CACGCCCAAT TTCITTAAGG 240  
 50 TACATACGAA CTGGGTCATT TATTTTAACA CCTGGAGGGG CACTAAGATC ACTTGGATTTC 300  
 AGTTTCTCGT CAGTATCTGA ACTATCTTTT TCATTAACTA GTGAAATATC ATTATCATTT 360

55

GCAATTTCTT CATGACTTAA ATGACCCTCT TTTTACCTT TTTCAATTAA TTGCTTCTTA 480  
 ACATCTTCTA ATGTTAATGT CGGATCAATT GTTGTGTTTT TAATTTTAAC TGTGTTATCA 540  
 5 GACATGAAAC GGCCTCCCGA TTTTAAATAT GAACATTCGA AATTTATTCA ATATTGCTAT 600  
 TTTAAACGAA ATTCTTAATT AATTCCATCC ATATTTTnAA TTTTATTTTA CAAATTGGGA 660  
 ACTAAATCCC CAATATTTAT TTTTCAATAG TGGTGGTT 698  
 10

## (2) INFORMATION FOR SEQ ID NO: 182:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5147 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

ACTTGATGAT GTATACAATG TATTTCAAGA ATATTATCAA AAAACATCTA ACATTAAGTT 60  
 TTGTAGAATT CACAATTCTA GCTATTATCA CTTCTCAAAA TAAAAACATC GTTCTTCTTA 120  
 25 AAGATTTAAT TGAAACAATC CACCATAAAT ACCCTCAAAC TGTTAGAGCT CTCAATAATT 180  
 TAAAAAAGCA AGGCTATCTA ATAAAAGAAC GCTCAACTGA AGATGAAAGA AAAATTTTAA 240  
 TTCATATGGA TGACGCGCAG CAAGACCATG CTGAACAATT ATTAGCTCAA GTGAATCAAT 300  
 30 TATTAGCAGA TAAAGATCAT TTACATCTTG TTTTTGAATA ATATCTCTAT TACGCAAGTG 360  
 TGCTGTATTC TAAAGTGCAC TTGTGTTTTT TATTTTTTAA TAAACCTCA GCACATAATG 420  
 AACAACTTTC TATTTTCTAT ATCACTTAAA ACCATTTCCG AAATTAAACC TCAGCACATT 480  
 35 CAAAGCCCCA CTTTATTCTT AAAAATATTT TTAACTCAT ATGTATTAAA CCGCTTTCAT 540  
 TATAAAAAAT ATCTCTATAT TtTATCTGtT TtTATTAATC GAAATAGCGT GATTTTGCGG 600  
 TTTTAAGCCT TTTACTTCCT GAATAAATCT TTCAGCAAAA TATTTATTTT ATAAGTTGTA 660  
 40 AAACCTACCT TTAAATTTAA TTATAAATAT AGATTTTAGT ATTGCAATAC ATAATTCGTT 720  
 ATATTATGAT GACTTTACAA ATACATACAG GGGGTATTAA TkTGAAAAAG AAAACATtT 780  
 ATTCAATTCG TAAACTAGGT GTAGGTATtG CATCTGTAAC TTTAGGTACA TTACTTATAT 840  
 45 CTGGTGGCGT AACACCTGCT GCAAAtgctG CGCAACACGA TGAAGCTCAA CAAAATGCTT 900  
 TTTATCAAGT CTTAAATATG CCTAACTTAA ATGCTGATCA ACGCAATGGT TTTATCCAAA 960  
 50 GCCTTAAAGA TGATCCAAGC CAAAGTGCTA ACGTTTTAGG TGAAGCTCAA AACTTAATG 1020  
 ACTCTCAAGC TCCAAAAGCT GATGCGCAAC AAAATAACTT CAACAAAGAT CAACAAAGCG 1080

55

	AAAGTCTTAA AGACGACCCA AGCCAAAGCA CTAACGTTTT AGGTGAAGCT AAAAAATTAA	1200
	ACGAATCTCA AGCACCGAAA GCTGATAACA ATTTCAACAA AGAACAACAA AATGCTTTCT	1260
5	ATGAAATCTT GAATATGCCT AACTTAAACG AAGAACAACG CAATGGTTTC ATCCAAAGCT	1320
	TAAAAGATGA CCCAAGCCAA AGTGCTAACC TATTGTCAGA AGCTAAAAAG TTAAATGAAT	1380
	CTCAAGCACC GAAAGCGGAT AACAAATTCA ACAAAGAACA ACAAATGCT TTCTATGAAA	1440
10	TCTTACATTT ACCTAACTTA AACGAAGAAC AACGCAATGG TTTTCATCCAA AGCCTAAAAG	1500
	ATGACCCAAG CCAAAGCGCT AACCTTTTAG CAGAAGCTAA AAAGCTAAAT GATGCTCAAG	1560
	CACCAAAAGC TGACAACAAA TTCAACAAAG AACACAAAA TGCTTTCTAT GAAATTTTAC	1620
15	ATTTACCTAA CTTAAGTGA GAACAACGTA ACGGCTTCAT CCAAAGCCTT AAAGACGATC	1680
	CTTCAGTGAG CAAAGAAATT TTAGCAGAAG CTAAAAAGCT AAACGATGCT CAAGCACCAA	1740
	AAGAGGAAGA CAATAACAAG CCTGGCAAAG AAGACAATA CAAGCCTGGC AAAGAAGACA	1800
20	ACAACAAGCC TGGTAAAGAA GACAACAACA AGCCTGGTAA AGAAGACAAC AACAAAGCCTG	1860
	GCAAAGAAGA CGGCAACAAG CCTGGTAAAG AAGACAACAA AAAACCTGGT AAAGAAGATG	1920
25	GCAACAAGCC TGGTAAAGAA GACAACAAAA AACCTGGTAA AGAAGACGGC AACAAAGCCTG	1980
	GCAAAGAAGA TGGCAACAAA CCTGGTAAAG AAGATGGTAA CGGAGTACAT GTCGTAAAC	2040
	CTGGTGATAC AGTAAATGAC ATTGCAAAAG CAAACGGCAC TACTGCTGAC AAAATTGCTG	2100
30	CAGATAACAA ATTAGCTGAT AAAACATGA TCAAACCTGG TCAAGAACTT GTTGTTGATA	2160
	AGAAGCAACC AGCAAACCAT GCAGATGCTA ACAAAGCTCA AGCATTACCA GAAACTGGTG	2220
	AAGAAAATCC ATTCATCGGT ACAACTGTAT TTGGTGGATT ATCATTAGCC TTAGGTGCAG	2280
35	CGTTATTAGC TGGACGTCGT CGCGAACTAT AAAACAAAC AATACACAAC GATAGATATC	2340
	ATTTTATCCA AACCAATTTT AACTTATATA CGTTGATTAA CACATTCTTA TTTGAAATGA	2400
	TAAGAATCAT CTAAATGCAC GAGCAACATC TTTTGTGCT CAGTGCATTT TTTATTTTAC	2460
40	TTACTTTTCT AAACAACCTC TGAAACGCCT CAACACTTTC TACTCTGATT ACATATATGA	2520
	CATTTTTAGG CATTAAAAA TCGAACTAGA CAAGATGCTC ATTGCATTTT GTACTAGTTC	2580
	GATTCATGAA TAATTAGATT TAAAATGTCA TTTGAATCCA AGTGACAACA TTATTTATAT	2640
45	TTAGAATATT AACGTTAGTA TAAACGTCCA AACACAAATA AAAGCAACAA ATATAATACT	2700
	GTATTTTAAC GTCATTTTAA ATAATGCAGA TTCTTCACCA ACTTTTTTAA CAGCTGCAGT	2760
50	CGCAATGGCA ATTGATTGTG GTGAAATAAG TTTCGCTGCT ACACCACCTG CAGTGTTAGC	2820
	TGCCACAAGT AATGAACCGC TTGTTGAAAT TTGTTGTGCC ACTGTGCTT GAATAGGTGC	2880

55



	TGGAGAGAAT AATGGGAAAA TTGCTCCCGC TTTAGCAATA CCTTGTCCAA TTGCTACAGT	3000
	CAAACCACCG TATGTCATAA CTTTAGCAAT AGCTAGGATA GCTGAAATTG TAAGGATCGG	3060
5	TAACCATAAT TCTTTAATTG CTTGACCAA TAAAGCACCT GCACTTTTCC ATTTTAACTT	3120
	CGTAATTAAA ATTGTAATAA TTACTGTAA TAAATCGCT GTCCCAGTTG CACCAATTAA	3180
	ATCGAGACGC AACGCAATTC CTTTAGGCGA TAAATCACTC ACAGTATTTG GAATTGGCAA	3240
10	TTTATTACT AAACCTTCAA GTGCACCTCC AGGTTGGAAT AATTTTTTGA AGAATGGTGC	3300
	ACTCCATACT AATACAAAGG CAGTTAAAAT TACGAACGGA CTCCAAGCAA AGACAATTTT	3360
15	TTTAGGCGTT CGTTTTTGAA TTTTATGTTT AGACGCTTCC AATCTGAAAA TGTTTTTTCGG	3420
	TTTAAATTTA CGACAAACAA ATGCTAACAC CACCATTGTT GCTAGTGATG GAATAATGTC	3480
	TGCTAGTTCT GGACCATGGA ATATTGTAA TAATAATTGT AATCCAGTAT ATGTACCACT	3540
20	CACTGTAAA ATGACAGGTA AAATTTCTTT AATACCTTTC ATACCATCTA CAATGAATAC	3600
	TAAACAAAT GGAATAATAA AGTTTAAAAT TGGAAGTGTT AATGCTGAGT ATCTCGCAAC	3660
	ATCTAATGTT GTAACGCCTC CACTTAAGTT AAACGTATCA ATAATACTAA CTGGTAAACC	3720
25	AATTGCACCA AAGGCACCCG CCGCACCATT AGCAATTAAA CATAACATCG CTGCTTTTAA	3780
	TGGTTCAAAT CCAAGTTGAA TTAATAATAC TGCACAAATC GCAATTGGCA CACCAAATCC	3840
30	TGCTGCACCT TCTAAAAATG CGTTGAAACA AAATCCAATT AATAATAGTT GGATTCCTTG	3900
	GTCCACTGAA ATACTTGCAA TACTATCTTG AATAATAGAA AATTGTCCTG TTTTAATAGA	3960
	AACTTTATAT AACCAAATG CCATTAAAAC GATATATCCT ATTGGGAAAA TACCGGCAAC	4020
35	AACGCCTTCT GTAATCGCAC CTGCTGATAC ACGCGCTGGT AATTCAAATA CAAATAAAGC	4080
	CACAATCAAT GTAACAACCA AAGTTGTCAA TGCTGCATAA ATGCCCTTCA TTTTAAAAAC	4140
	GGTTAAGCAT AATAAAAATA AAATAATAGG TACTGCTGCA ACTAAGGCTG ATAATCCGAC	4200
40	ATTATCGAAT GGATTACAG TAAGTAGTGT CATAATGACT CCCTCTCTT ATATAAATA	4260
	TTTATCATTC TGATTAACTT ACAACCTATT TCAACTTATA TTTTGCATG ATCACATATT	4320
	TAAATGTAA CACTCCTATA TGTGACAGGC AATCGAATTT TTACAAAAAG TTCACAAAT	4380
45	ATACACAATA TTTAACTATA ATAmATAATA TATCaTnCTA ATTATAAATA CTAGATATTA	4440
	TTTATAATAA TCTCAGGAAT TCGCTTCAA ACTGCATCAT GAGAGTTTAT ATTTTTATTG	4500
50	AGAATCTCTC ATTTTATGAA TTGTAGGAAG TAAACAAAAT ATGACAAGCG TCAAACCAAT	4560
	GATAATGATA AATATCATAT TAAACCATAG TAAATTGAAT TGATGATGGT GTTGTATTTG	4620
55	CCAAATTTCT AATACTGTGA AGATAGACAT ATAGCTCATA ATCTCTAAAT TTAACGTACT	4680

AAATCGTTCA TAGTATCTAC CTGCAATGAA AAATATAAGC CAAATCACTA TAAATGCGCT 4800  
 ATTAATCAAA AGCAGCACCC ATTTATCAGC AAAATTATCA GCATCCCCTG CTAAATTATA 4860  
 5 ATGAATAGGC ACTTTGGTTG GTAATTTTGG ATAGGTCACT ACTGTATAGC ACATCATAGC 4920  
 TAAGTAAATA AGTAGACTTA ATATTGTAAA AGACCTGATT TTAGACATTC TATCGCCTcT 4980  
 TcTTTACATT TTATGTATAA CACTCTGCCT ATTTTACCTT TTAATaCATT ACCCCAACGA 5040  
 10 TtAAaCAATA tGTAAaTGATA CTATAATTGC GTCAGGAGTA TCCGCTTGTT AAATGTGCAT 5100  
 AGCTTATATT TAGCTGTTTA ACATGCCACA TAATGATTcG AATTATT 5147

15 (2) INFORMATION FOR SEQ ID NO: 183:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1312 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

25 CACTTACTTC CACCATTATC ATAACCTTAA AATGGATATA nTTCATCAAA CATTATCTAA 60  
 AGGCGTCGCA CCTACACCAA CACCATCCAA CAATTAACTT ACAACTCTGC GATTACTTCT 120  
 TCAGCAGCAA CTTTCACnTG CGTAATACAA TCAGGTAGTC CAACCGCTTC AAAAGATGCA 180  
 30 CCAGTTACTC TAAGTCGTGG ATATGTTTGT TTAATATGTG CTTGAATCTG TCTAATTTGT 240  
 TGAATATGAC CGACATGGTA CTGTGGCATA CTTTTCGGCA AACGATTGAC AATTGTAAAT 300  
 TCAGGATCAC CTTTAAATGT CATCATTTGA CTTAAATCTC TACGTACAAT CGATACTAAT 360  
 TCATTATCTG TATGATCATC AACCACAGTA TCACCTGGTT TACCTACATA CGCACGAATC 420  
 AAAACCTTAC CTTCCGGTGT AGTAAATGGC CATTTTTTCG ATGTCCAAGT ACATGCGGTA 480  
 40 ATGTCTGTAT CACTCGTTCT CGCAATTACG AAGCCAGTAC CATCATGGGT ATTTTCAATG 540  
 TCTTTTTCAT CAAATGCCAA TACAACAGTT GCAACAGTCG TACTATCCAT CGTTTTAAAG 600  
 TAATCAAATG CTGGATCTTG TCCGAACCAA TTTAAAAACA CTTGATGTGG TGTCGTTACT 660  
 45 AATACGCCAT CATACTTTC TTCTAGTTGA TCATTGTAAA CAATTTTATA TTGTTTTTGA 720  
 GATGTAATTA TATCATCCAC TGACGTATTG TAGCGTATTG TCACACCTTT ATTTTAAACA 780  
 TCTTGTTCTA ATGCTTCAAT AAATGAGCTT AAACCATGCT TAAATTGTTT GAATTGTCCT 840  
 50 TTCGGTGCGC CAGGATATAA TTGTCTTTGT TTCAGACGCT TATTTTTCTC ATCCTTCATA 900  
 CCTTTTATCA GACTTCCGAA TGCCTCTTCT TTTTCTTTAA AATTAGGAAA CGTACTCATC 960

55

TCAAGTACCT CATTACCTAA TCTTGCTCTG AAAAATGCAC CAACAGAAAT GTCACCATCC 1080  
 TGCATTGAG TAGGTTTTTT TAATAAATCA AACCTGCTC TTAATTTACC AAGTGGCGAT 1140  
 5 ATTAATTTTG TAGTAACAAA TGGTTTAATA TCTGTTGAA TACCCATAAT TGAACACCT 1200  
 GGAATCGGAT ATAATTTATT TTTCGCAAAA ATATATGATT GTCCAGTCGT ATTTGTAACA 1260  
 ATATCTTGTT CTAATCCAAT ATCTTTCGCT AATTCTGTCA TAATCGTTTT TC 1312  
 10 (2) INFORMATION FOR SEQ ID NO: 184:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6157 base pairs  
 15 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:  
 20 TTTTACAATA AAAATATGAT ATACTACTTG TCGTATATAA GGAACGGAGG ACAATTTATG 60  
 CATACTTTT TAATCGTATT ATTAATCATT GATTGTATTG CATTAAATAAC TGTTGTACTA 120  
 25 CTCCAAGAAG GTAAAAGCAG TGGACTTTCA GGTGCCATCA GTGGTGGTGC TGAGCAGTTA 180  
 TTCGGTAAAC AAAACAACG TGGCGTCGAT TTATTCTTAA ATAGATTAAAC AATTATTTTA 240  
 TCAATATTAT TTTTGTACT TATGATTTGC ATAAGTTATC TTGGTATGTA AGGTCCGGCG 300  
 30 ATGTAAATGT CGGGCTTTTT TATTTATAAT TAAGAATGTA ATAGTTTAAAC AATAAGCTAT 360  
 GTAAAATATA TAGCCTAGTT AAGTATGCAA AGGGAGCGTT AGATTTATGC AGATAAAATT 420  
 35 ACCAAAACCT TTCTTTTTTG AGGAAGGTAA ACGTGCCGTG TTATTACTAC ATGGTTTTAC 480  
 AGGCAATTTC TCTGATGTTG GTCAATTAGG TCGATTTTAA CAAAAGAAAG GTTATACATC 540  
 ATATGCACCG CAATATGAAG GCCACGCGGC ACCACCAGAT GAAATACTGA AATCTAGTCC 600  
 40 TTTCTGTTGG TTAAAGATG CGTTAGATGG TTATGATTAT CTTGTTGAAC AAGGTTATGA 660  
 TGAAATTGTT GTTGCTGGTC TATCATTAGG TGGGGATTTT GCTTTAAAT TAAGCTTAA 720  
 TAGAGATGTA AAGGGTATTG TAACGATGTG TGCTCCTATG GGTGGCAAAA CTGAAGGTGC 780  
 45 CATTTATGAA GGCTTTTTAG AATATGCACG CAATTTTAA AAGTATGAAG GTAAAGATCA 840  
 AGAGACTATT GATAATGAAA TGGATCATTT TAAACCAACT GAAACTTTAA AAGAATAAG 900  
 TGAAGCATTG GATACGATTA AAGAGCAAGT TGATGAAGTG TTGGATCCTA TTTTAGTGAT 960  
 50 TCAAGCAGAA AACGACAATA TGATTGATCC ACAATCCGCA AATTATATAT ATGACCATGT 1020  
 AGATTCTGAT GACAAAAATA TCAAGTGGTA CAGTGAATCT GGACATGTTA TTACGATTGA 1080  
 55

	AGAATAAAAA GAGATTTTAA CATTAGAAAAG GAGGGGCATA ATGAATTTAA AGCAATCTAT	1200
	AGAAGAGATT ATTAATCAAC CTGAATATGA ACCTATGTCA GTGTCAGATT TTCAAGATGC	1260
5	ATTAGGTTTA AGCAGTGCCG ACTCGTTTAA AGATTTAATT AAGGTGCTTG TGGAGTTAGA	1320
	ACAATCAGGA TTAATCGAAC GTACAAAAAC AGACAGATAC CAAAAAAGC ATAGTTATAG	1380
	AGGTCAATCA AAATTGATAA AAGGAACGTT AAGTCAAAAT AAAAAAGGCT TTGCATTCTT	1440
10	AAGACCTGAA GATGAGGATA TGGAAGATAT ATTTATTCCC CCGACGAAAA TTAATCGTGC	1500
	CTTGGATGGA GATACTGTTA TTGTAGAAAT CCATCAATCA AAAGGTGAAC ATAAAGGTAA	1560
15	AATCGAAGGG GAAGTTAAGT CGATTGAGAA GCATTCTGTA ACTCAAGTTG TTGGTACGTA	1620
	TAGTGAAGCT AGACATTTTG GCTTTGTTAT TCCGGATGAT AAACGTATTA TGCAAGATAT	1680
	TTTCATTCCCT AAAGGTCAAA GTTTAGGCGC AGTCGATGGT CATAAGGTAC TTGTACAAAT	1740
20	TACTAAGTAT GCTGATGGTT CAGATAATCC AGAAGGACAT ATTTCTGCTA TTTTAGGACA	1800
	TAAAAATGAT CCTGGCGTAG ATATTTTATC TATTATCTAT CAACATGGCA TAGAAATTGA	1860
	ATTCCTGAT GAAGTGTTAC AAGAAGCTGA AGCAGTACCT GATCATATTG AAAATACTGA	1920
25	AATTAAAGGC CGTCATGATT TACGTGATGA ATTGACAATC ACAATTGATG GTGCTGATGC	1980
	TAAAGACTTA GATGACGCAA TTAGTGTTAA AAAGTTAGCG AACGGTAATA CGCAATTAAC	2040
30	TGTAAGTATT GCTGATGTCA GCTATTATGT AACAGAAGGT TCTGCATTGG ATAAAGAGGC	2100
	ATATGATAGA GCGACAAGTG TATATCTTGT TGACCGTGTA ATTCCAATGA TTCCACATCG	2160
	ATTAAGTAAT GGTATTTGTT CATTGAATCC TAATGTTGAT CGTTTAACTC TAAGCTGTCTG	2220
35	CATGGAAATC GATGCTAGTG GTCGCGTTGT TAAACATGAA ATTTTGTATA GTGTTATACA	2280
	TTCTGATTAT CGAATGACGT ATGATGCGGT AAATCAGATT ATTACTGAAA AGGATCCTAA	2340
	CATTGCGGAA CAATATAATG AAATTACGCC TATGCTAGAT TTAGCACAAG ATTTATCTAA	2400
40	TCGTTTGATT CAAATGAGAA AACGACGTGG TGAAATCGAT TTTGATATTA GTGAAGCAAA	2460
	AGTATTAGTT AACGAAGACG GTATACCAAC AGATGTTCAA TTAAGACAAC GTGGCGAGGG	2520
45	TGAACGTCTA ATTGAATCAT TTATGTTAAT TGCAAATGAA ACAGTTGCTG AACATTTTAG	2580
	TAAGTTAGAT GTACCTTTTA TTTACCGAGT GCATGAGCAA CCTAAATCAG ATCGCTTAAG	2640
	ACAATTCTTT GATTTTATTA CAAACTTTGG CATCATGATT AAGGGTACTG GCGAAGATAT	2700
50	TCATCCAACA ACACTTCAAA AGGTTCAAGA AGAAGTAGAA GGTGACCTG AACAAATGGT	2760
	CATTTCAACA ATGATGTTGC GTTCAATGCA ACAAGCGCAT TATGATGATG TGAACCTGGG	2820
	ACATTTTGGC TTATCAGCTG AATATTATAC GCATTTTACA TCACCAATTA GACGTTATCC	2880
55		

EP 0 786 519 A2

	AGAAGTGAAG	CGTTGGGAAG	ACAAATTGCC	TGAGTTAGCT	GAACATACTT	CTAAACGTGA	3000
	ACGTCGTGCT	ATTGAGGCAG	AACGTGATAC	TGATGaATTG	AAAAAAGCAG	AATATATGAT	3060
5	TCAACATATT	GGTGATGAAT	TTGAAGGTAT	TGTCAGCTCA	GTAGCTAACT	TCGGTATGTT	3120
	CATTGAATTG	CCAAATACGA	TAGAAGGTAT	GGTTCATATT	GCGAATATGA	CTGATGATTA	3180
	TTACCGTTTT	GAAGAGCGTC	AAATGGCATT	AATTGGTGAG	CGTCAAGCTA	AAGTATTTAG	3240
10	AATTGGTGAC	ACAGTTAAGG	TTAAAGTGAC	GCATGTTGAT	GTAGATGAAC	GATTAATTGA	3300
	TTTTCAAATT	GTAGGTATGC	CTTTACCGAA	AAATGATCGA	TCACAGCGCC	CAGCGCGAGG	3360
	TAAGACAATT	CAAGCCAAAA	CGCGTGGTAA	ATCATTAGAT	AAATCAAAAT	CTGATGATAA	3420
15	GGGTCGTAAG	AAAAAAGGTA	AGCAACGTAA	AGGTAAAAAC	CAACGTAATA	ATGATAAATC	3480
	AGGTAATAGT	AAGCATAAGC	CATTTTATAA	AGATAAAAGT	GTGAAAAAGA	AAGCACGTCG	3540
20	TAAGAAAAAA	TAAGCAGCAA	TGAGGTGAGT	ATGAATGGCT	AAGAAGAAAT	CACCAGGTAC	3600
	ATTAGCGGAA	AATCGTAAGG	CAAGACATGA	TTATAATATT	GAAGATACGA	TTGAAGCGGG	3660
	AATTGTATTG	CAAGGCACAG	AAATAAAATC	AATTCGCCGA	GGTAGTGCTA	ACCTTAAAGA	3720
25	TAGTTATGCG	CAAGTTAAAA	ACGGTGAAAT	GTATTTGAAT	AATATGCATA	TAGCACCATA	3780
	CGAAGAAGGG	AATCGTTTTA	ATCACGATCC	TCTTCGTTCT	CGAAAATTAT	TATTGCACAA	3840
	GCGTGAAATC	ATTAAATTGG	GTGATCAAAC	ACGTGAGATT	GGTTATTCGA	TTGTGCCGTT	3900
30	AAAGCTTTAT	TTGAAGCATG	GACATTGTAA	AGTATTACTT	GGTGTTGCAC	GAGGTAAGAA	3960
	AAAATATGAT	AAACGTCAAG	CTTTGAAAGA	AAAAGCAGTC	AAACGAGATG	TTGCGCGCGA	4020
35	TATGAAAGCC	CGTTATTAAG	CGATTTAGTT	GCTTAATCGG	GCTATATTTG	ATATAGTTAT	4080
	ATGTGCTTTT	GTAAATTACA	AAAGTATGAT	TTGTTTGATT	TATTATTTTCG	GGGACGTTCA	4140
	TGGAFTCGAC	AGGGGTCCCC	CGAGCTCATT	AAGCGTGTCT	GAGGGTTGTC	TTCGTCAATCA	4200
40	ACACACACAG	TTTATAATAA	CTGGCAAATC	AAACAATAAT	TTCGCAGTAG	CTGCCTAATC	4260
	GCACTCTGCA	TCGCCTAACA	GCATTTCCCTA	TGTGCTGTTA	ACGCGATTCA	ACCTTAATAG	4320
	GATATGCTAA	ACACTGCCGT	TTGAAGTCTG	TTTAGAAGAA	ACTTAATCAA	ACTAGCATCA	4380
45	TGTTGGTTGT	TTATCACTTT	TCATGATGCG	AAACCTATCG	ATAAACTACA	CACGTAGAAA	4440
	GATGTGTATC	AGGACCTTTG	GACGCGGGTT	CAAATCCCGC	CGTCTCCATA	TTTGTAGCCT	4500
50	ACAGCCTTTG	TGGTTGTGGG	CTTTTTTATT	TTGTGTTTTT	CAGGGGATAA	TGCATTGCAG	4560
	AATTTGTTGT	GAGTATTGAT	ATAGCAGTGT	TTGTATAGGT	GTTTATTTGA	TGGAGGAAAG	4620
	AGTAATAAGT	GATTATGAAT	TAGTTTTTGA	GATATAAGGG	GACAGTGATG	TGTGTCAAAT	4680
55							

TTATACGCAA AAAATTCTCC ATGTTATATA TGTCAATATA AAAATGTGAA TCGTCTACAC 4800  
 TTAATTGGAT AAATGGCTAC TGAAAAAGAA CTTTTCATTT TTGTTACGTC ACTAAGTGGG 4860  
 5 TGTAGTTATA AAGAGATGAG CCGAGTTTTG ATATTTTCAT TAGAATCAAT ATGCCTATTA 4920  
 ACACAATCAG CAATAGTTGA CGAGACGGAA ATAAAAGAAG TCGTAGTTAA GAAATGCATT 4980  
 TCACAACATA CCATTGTAGC CATTTTTATT GTTTTGGATG ATAAACTCTT TTTGGAATTT 5040  
 10 TTAGTTTTTA TAATTTGCAA CTACACTACT TCTTTTACTA ATATTAATGT CTAAGTAATC 5100  
 GATAAAAAAT TTTCCATTGA ATAAATGAGA AGTTAAAAAC TTTACTTAAC CTTTCyCATT 5160  
 15 GCATTTTCCT ATTCACGATT TTAAGAACCC AACATACTAC AAACGAATTT TAAAAGGCGA 5220  
 GAGTAAAGCT TACTTGTTTA TTATACATAT TTAAAATCCA AGAGTCAGAA CAGACTACTC 5280  
 CTCTTTATAA CTATAAAAAA TAGCTATGAA AAAATCTATC GTCATAGATT CCTTCATAGC 5340  
 20 TAATCTTAGT ATGTTTATTT TTATTTTAGG ATGCTATTTA TCAACTCAAC ATATAACTCA 5400  
 CTATTTTTAT AACCTTCTAA TATATCATT ACTTGTCTAA TAGGTATTTT TGGTACTTCT 5460  
 CTAATGTTTT CCAATTTTGT TTTAAATTGT TTTTTGTTA TTTGCTCTTT ATTTGTAGCC 5520  
 25 AATTGGAACA AGTAAGAATC TAGCATATTA ATTTCTTTAT ATGAATACAT ATATCTTAAT 5580  
 AACACTAAAT CTCTAGTTTT TAAGTTAGGC GCTAGTTCTT CTGTGAATTG TTCTATTGAT 5640  
 TGTyTCATTA ATAACAATCT CATTTCTAAT TCTTCATTAT TCATTTTATC AACTCTTTtT 5700  
 30 TATATTAATG CTGACCAAC TTGGGAAACC CAAAACCTA TGCTTCTTGC AGTAGAATCT 5760  
 TTAATACCAG TTCCCATCAA TGCTTGTAAG ACTTGACCTT GTACATTTCC CCATGTAGCC 5820  
 35 TCTTCTTGTT TTAATGCATT ATTCAATGCG GGATTTACAA ATTTATCCCA TCTTTTTTTT 5880  
 ATGATTTTCC GGCACGGGGA CTGATTTCTT TAACACCATT AAACACAGAT TTTTATTTT 5940  
 TAATCATAGC TTTATAGTAT CATGTTGGCT AAGCTATAAA TAAGTCAGTT TCTCTAAAAA 6000  
 40 TTAAATAACT GAATGTAAGA CAATCAACAA wCCAAATTTA TACTTCATCT AAACCACTGT 6060  
 GGTGTCATC TTTTGTCTTT TCTTTTCTT TCTCTCGTTC TTGTTCTTTT TTGTACTCTT 6120  
 CTTCAAATTC TTTTCTTTT TTTTCTACTT CTTCTCT 6157

(2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 884 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

CATTGTGTAT TCTGAGTAGC CAATTTGGCA AAGATGAACA AACGTCTGAA CAAACGTATC 60  
 AAGTTGCAGT CGCATTAGAG TTAATTCATA TGGCAACACT TGTTTCATGAT GACGTTATTG 120  
 5 ATAAAAGCGA CAAGCGTCGA GGCAAGTTAA CCATATCAAA GAAATGGGAT CAGACAACCTG 180  
 CTATTTTAAC TGGGAATTTT TTATTGGCAT TAGGACTTGA ACACTTAATG GCCGTTAAAG 240  
 ATAATCGTGT ACATCAATTG ATATCTGAAT CTATCGTTGA TGTTTGTAGA GGGGAACTTT 300  
 10 TCCAATTTCA AGACCAATTT AACAGTCAAC AGACAATTAT TAATTATTTA CGACGTATCA 360  
 ATCGCAAAAC AGCACTGTTA ATTCAAATAT CAACTGAAGT TGGTGCAATT ACTTCTCAAT 420  
 CTGATAAAGA GACTGTACGA AAATTGAAAA TGATTGGTCA TTATATAGGT ATGAGCTTCC 480  
 15 AAATCATTGA TGATGTATTA GACTTCACAA GTACCGAAAA GAAATTAGGT AAGCCGGTCG 540  
 GAAGTGATTT GCTTAATGGT CATATTACGT TACCGATTtTT ATTAGAAATG CGTAAAAATC 600  
 CAGACTTCAA ATTGAAAATC GAACAGTTAC GTCGTGATAG TGAACGCAAA GAATTTGAAG 660  
 AATGTATCCA AATCATTAGA AAATCTGACA GCATCGATGA GGCTAAGGCA GTAAGTTCGA 720  
 AGTATTTAAG TAAAGCyTTG AATTTGATTT CyGaGTTACC aGATGGACaT CCGaGAtCAC 780  
 25 TACyTTTAAG TTTGACGAAA AAAATGGGTT CAAnAAACAC GTAGTATTTA TGnAAAAGTA 840  
 TTGAAAGCGC TTTACCAACC TGTTAATATA TAATAGTAAT ATAC 884

(2) INFORMATION FOR SEQ ID NO: 186:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6876 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:  
 40 AATTTTCATCT GTCGTGCAA AATCTTTGTT TTTCCTTGCT TCATTACGCT CTTGATTAA 60  
 TTTTTCACAA TCTTCATCCA ATAATTCATC TGCATTTTTA GATTTTAACG GTACACCTAA 120  
 AACATCGCTG AAAATTTGAT AAAGTCTTTT AAATTTATCA ATTACTTCTG TTGATGTTGT 180  
 45 GTTCTCTAGT ACATATTTAT TCGCAAGTtT TGCTAAATCA TACCAAGCTG TAATTGCATT 240  
 AGCTGTATTA AAATCATCAT TCATAACTGT TTCAAAACGA TTTAAAATCG CATCAATTTG 300  
 ATCAATATAT GTCTGTTGAT TTTCAATATT AGTAGCAATT TGTGCGCGCT CTTCAATTAA 360  
 50 TTGATAACTA TTGCGAATAC GCTCTAGTcC aCTACGTGCT GATTCTACCA ATTCTAGATT 420  
 ATAGTTAATT GGGCTTCTAT AATGTACGCT AATCATAAAG AATCTTAGTA CATCTGGATC 480

55

	ATTATCAATA TTAATGAAAC CATTATGCAT CCAATAATTA GCAAATGGCG CATGATTATG	600
	TGCTTCTGAT TGTGCTATTT CATTTTCATG ATGTGGAAAT TGTAATCTG AACCACCCGC	660
5	ATGTATATCA ATTGTAGGTC CTAGCTCATG AAATGCCATT ACAGAACATT CTATATGCCA	720
	TCCTGGTCTA CCTTCACCAA ATGGGCTATC CCAACTAATC TCGCCAGGTt CGCTTTTTTC	780
10	CACAATGTAA AATCAAGTGC ATCTTCTTTA TGCTCTCCTG CATCTATACG AGCACCCACT	840
	TTTAAGTCAT CTATGGATTG ATGACTTAAT TTACCATAAC CTTCAAATTT ACGTGTTCTA	900
	AAGTAAACAT CGCCACCACT TTCATATGCA TAACCTTGAT CCACCAAATC TTTAATAAAT	960
15	TGAATAATGT CATCCATATG GTCCATTACC CTTGGATTTC AAGTCGCTTT TCTAACATTT	1020
	AACGCACCAA CATCTTCATG AAAAGCAGCG ATATATTTTT CTGCAATTTT GGAACAGAC	1080
	TGATTTAATT CTTGAGAACG TTTAATTAAT TTATCATCTA CGTCTGTAAA ATTTGATACA	1140
20	TATTCTACAT TATATCCTTG GTATTCAAAG TAACGTCTCA CTACGTCATA ATTAATTGCW	1200
	GGTCTTGCGT TACCAATATG AATGTAGTTA TATACAGTAG GACCACATAC ATACATTTTT	1260
	ACTTCCCTG GTTCTATAGG CTTGAACACT TCTTTTTGAC GTGTAAGCGT ATTATATAAT	1320
25	GTAATCATCT TGAATCTCTC CATTCCTAGT CTTTTCAAGT TGTCGTTCTA AATGCTTAAT	1380
	TTGTTCATAA ATTGGATCAG GTAGATGGCG ATGATCAAAT GTTTTTCCAA CTCGAACACC	1440
30	ATCTTGCTTA ACAATATGTC CTGGTATACC AACAACCGTT GAATAACTTG GAACTGATTG	1500
	TAAAACAACT GAAATTTGCAC CAATATTTAC ATTTGAATTT ATTTTAATAT TTCCTAAAAC	1560
	TTTCGCACCG GCTGCTATTA AAACATTGTC TCCTATATCT GGGTGTCTTT TCCCTCTTTC	1620
35	TTTCCCTGTC CCACCAAGTG TCACGCCTTG ATAGATTGTC ACATTATCAC CAATTGTACA	1680
	TGTTTTCTCCT ATTACAACGC CCATACCATG ATCTATAAAT AGACGCTTTC CAATTTTAGC	1740
	ACCTGGATGG ATTTCTATAC CTGTGAAAAA TCTTGAAATT TGAGATATCG CGCGTGCTGC	1800
40	AACATATTTT TTTTGGTTGT ATAACCTATG TGCAATCAAA TGACTCCAAA CTGCATGTAA	1860
	ACCTGCATAC GTTGTAATGA CTTCTAATGT TGAACGTGCC GCTGGATCCT GCTCAAATAC	1920
	CATTTTTATA TCGTCTCTCA TTCTTTTAA CAAGATCATT TCCTCCTCAA TGATTGAACT	1980
45	ACGTAAATAC ATAATTGAAG TACCTGCGAA ATTAAATATC AAAAAAGCAC CACTAACATA	2040
	CAAATTGTAT TGTTAGAGGC GCTCCGCAC GGTCCACTC TGAATTTAGC GAATAACATT	2100
50	AATAATATTG CGGGCGCTTC CAAATTATCA AGGAAACTAA GTCAACTTAA TGCTCATCAC	2160
	TCTCATTATA TATTTAATTC ATTTTACGAA GGTGCATTCA TTAATTTCTA CGTTGTACTC	2220
	ACAGCAACCG TACACTCTCT GCATCGTATA AATTTAATTA CTAATCCTTC GTTTTATATA	2280
55		



	ATAAAATTCA AGTATATACT ACCTTGATCT TGTCTATTTT ATTACTTATA TTGTTTTTAAA	2400
	CGGTTTACGA CTTTTTCTTT ACCAAGTACT TCAATTGTAT TTGGTAATTC AGGACCATGC	2460
5	ATTTGGCCTG TTACAGCAAC ACGAATAGGC ATAAATAATT GCTTGCCTTT TATTCCTGTT	2520
	TCTTTTTGAA CTTCTTTAAT TGTCTTTTTA ATTTAGCCG CTTCAAATGG TTCAAGTGCT	2580
	TCTAATTTAC TGAATAAGTG CGTCATTAAC TCTGGTACTT GCTCTCCATT AATCACTTGT	2640
10	TGTTCTTCTT CACCAAGAGC TGGCATTCTT TTAAAGAACA TTTCTGATAA AGGTACAATT	2700
	TCACCGGCAT AACTCATTTT TTTTGTATA AGCGCAATTA ATTTGCGTCC CCAAGATAAA	2760
15	TCCTCTCTG ACGGCACCTC AGGAATCAAA TTTGCTTTAA TTAAATGAGG TAATGCTAAT	2820
	TGGAATACTG TTTCAGTATC TTTTGTGTTT ATATATTGGT TATTAACCCA TGCTAATTTT	2880
	TGCTTATCGA AAAATGCTGG TGATTTTGAC AAACGCTTTT CATCAAAGAT TTTGATAAAT	2940
20	TCTTCTTTAG AAAAGATTTT TTCTTCACCT TCAGGAGACC AACCTAATAA CGCAATAAAA	3000
	TTAAATAACG CTTTCAGGTAA ATAACCTAAG TCACGATATT GCTCAATAAA TTGTAAAATT	3060
	TGCCCATCAC GTTTACTTAA CTTTTTACGT TCTTCATTAA CAATTAATGA CATATGACCA	3120
25	AAACGAGGTG GCTCCAGCC AAATGCTTCA TAAATCATAA TTTGTTTAGG CGTGTGTTGAA	3180
	ATATGATCAT CACCACGAAT TACATCTGAA ATTTGCATGT AATGATCATC TATAGCTACT	3240
	GCAAAATTGT ACGTTGGAAT GCCATCTTTT TTTACGATAA CCCAGTCACC AATACCATTT	3300
30	GAATCAAATG AAATATTTCC TTTTACCATA TCATCAAATG AATACGTTTG GTTTTGAGGT	3360
	ACTCGGAAAC GAATTGATGG TTGGCGTCCT TCTGCTCAA ATTGTTGACG TTGTCTTCA	3420
35	GTCAAATGCG CATGTTGACC ACCATAGCGA GGCATTTTAC CACGAGCGAT TTGCGCTTCA	3480
	CGTTCAGCTT CTAATTCTTC TTCTGTCATA TAGCATTTAT ATGCTTTATC TTCTGCTAGT	3540
	AACTGATCTA TTAATGGTTG GTAGATATGT TGACGTTTAC ATTGACGATA TGGTCCGTAG	3600
40	CCATTGTCTT TATCTACAGA CTCATCCCAA TCTAATCCTA ACCATTTAAG ATTATCAAAT	3660
	TGTGATGTTT CTCCATCTTC TAAATTACGT TTTTATCAG TATCTTCAAT TCGAATCACA	3720
	AAATCTCCGT TGTAATGTTT AGCATACAAG TAATTGAATA ATGCTGTTCT TGCATTACCA	3780
45	ATATGAAGAT ACCCAGTTGG ACTTGGTGCA TATCTTACTC TTATACGATC GCTCATTTTT	3840
	TTCACTCCTA AATTAAATAT CAGATTTTCA AGTTAGTTCA TATAAATTGT TCATTTGCTA	3900
50	TCTTCGACCG TCATAACAAA TGTCTAACTC GTCTTATTGT TAAAACGAAA CAATGCTTTT	3960
	TAACATGACC TTAATAAAT TTCAATGTTT AATCATAACA TAATCCCTG GGTAATATGC	4020
	TTAAATTTTA AATAGAAAGC TGTTGTTTTT TCAACACTTT AAAAAAGCTA TCCCTAAGAA	4080
55		

	TTAAACTTCA AATTAACTAT TCAAATACGT TAAAATTGAT TCTAATTTTG TATGTCTTGA	4200
	TTGCTATAAG AATAACTTTA TTAATATCTA AAATTTAACA CTTAATGAAC TTGTTTCAAT	4260
5	GATATATTAG CACTATTTGT ATTTTTTGAT AACTAATATG TTTTGCATTT ATTTATAGTT	4320
	ATACTTCAAA TTACAAACTt CGCCATTTCA TATACCTTTT AATATCTATT TTGTTTTCGT	4380
10	CAACTACAGT TTTTATAATG ATACTGTATC TTCGATTTTT TTAGCAAAAA CAATTCTTCC	4440
	TGAAGATGTT TGCAATAAGC TGACTACTTC TAAATTGACA TGACTGCCAA TAAGATTTTT	4500
	AGCATTATCA ACAACTACCA TCGTACCATC ATCTAGATAT CCTACTGCCT GACCAGGctC	4560
15	CTTACCCATT TTTGTCAGTA AAATATGCAG TTGATCACCT TGATGTACAT TAGGTTTGAT	4620
	TGCTTCTGAT AAATCATTAA CATTTAATGC TTTGATACCA TGTACATGAC AAACTTTATT	4680
	TAGGTTGAAA TCTGTCGTTA TAATACTTGC ATGATATTGT TTTGCTAATT TTAATAACAT	4740
20	CGTATCAATA TCACTATGTG TTTTAGTTGG ATGTATAACC TTTGTAGGAT AGTCTAAATC	4800
	ATACAATTCA TTTAAAATAT CTAAGCCTCT TTTACCCTTT TCaCGTTTAA CACTGTCATT	4860
	TGAATCTGCA ACAATTTGTA ATTCAATTAAT AACACCTTGT GGAATTAAAA TATTGCCATC	4920
25	GATAAAACCG CAACGAATGA CTTCTAAAAT ACGACCATCA ATAATTGCGC TTGTGTCGAT	4980
	AATTTTTGGC GTAgcaCTTT TaGTATGTTG TGACATGGAA CGCGCTATAT TCTCAGGTAA	5040
30	AAACATTAAC ATTTTCATCTC GTTTTTTAAG GCCAAATTGG AAACCGAAAT AACATAGTAA	5100
	TATCGTAATT ATGACAGGAA TGAAATGATT AAAAATAGAG TTGCCAATTG ATTCTAATAT	5160
	AAACGACACC ATAACAGAAA TAAGTAATCC GATTATTAAA CCTATTGTTG CGAATAGTAT	5220
35	TTCAACAGCA CTTCTACGCA TAATAAAATG TTCTAAACCT TTTATAGCGT TAGTAACTCG	5280
	TCTAATAAAT ACACCAAAAA TTAAGAACAT AAAAATACTA CCGATAATGC CATCTACATA	5340
	GTGATTTTTT AAAAAGCTGG AGTTTTGTAA TCCAAGATCA TTTGCAATT T CAGGAATAAT	5400
40	AATTATTCTT AATGCGCTCC CAATAATTAA GTAAATAATA ATAACCATTA GTTTAACGAT	5460
	ATTCACACAA TGTCCTCCTT TCTTGATGTT TTATGAATGA AGAGCAAATG ACAATACTTC	5520
	ATGTACAGTA GTTACACCTA TTACTTGAT ACCTTCAGGA TATGTCCATC CGCCTATATT	5580
45	ATTTTTAGGA ATAATTACAC GTTTGAAACC TAGTTTTGCA GCCTCTTGCA CGCGTTGTTC	5640
	TATCCGAGAT ACACGACGTA CCTCACCGT TAAACCAACT TCTCCAATAT AGCAATCTAA	5700
50	TCCGTCGACA GCTTTATCTT TAAAGCTAGA TGCAGTTGCT ACAATTACAC TTAAATCAAC	5760
	TGCTGGCTCC GTTAACTTTA CACCGCCAGC TACTTTGATA TAAGCATCTT GTTGTGTGTA	5820
55	TAGATAATTT TCTTCTTTT CCAAAACAGC CATCAACAAA CTTAATCGAT TATGATCAAT	5880

TATTAAAAGT GGTCTGGTTC CCTCCATGGT TGCAACAATT GTTGAACCTG GAACATTTGT 6000  
 TGAACGTTCT TCTAAAAACA TTTCAGATGG ATTATTTACA CCTTTTAATC CACTTTGCTT 6060  
 5 CATTTCGAAG ATTCCcATTt CATTcGTtGA ACCAAAACGG TTTTTAACAG CTCGCAAAAT 6120  
 TCGATATGCG TGGTGTTCAT CGCCTTCAAA ATAAAGCACA GTATCaACCA TGTGTtCTAG 6180  
 CAATCTTGGG cCCAGCAATT TGACCTTCTT TCGTTACATG ACCCACTATA AAAGTTGCaA 6240  
 10 TGTTcATTtG TTTAGCAATA TTCATTAAAC TTTGTGTACT TTCACGAACT TGTGAAACAG 6300  
 AACCTGGCGC AGAGCTGATT TCAGGATGAT ATATTGTTTG AATCGAATCC ACTACTAATA 6360  
 15 AATCAGGTTG TTCTTCTTTT ACTGTTTGAT AAATAACTTC AAGATCTGTT TCAGCTAATA 6420  
 CTTGCAATTC ACTTGAATCT TCATCTAATC GCTCTGCACG TAATTTAGTC TGACTAAGCG 6480  
 ATTCTTCTCC AGTAATATAT AGTACTTTTT TCTTTTGAGA TAACGATGCA CAAATTtGTA 6540  
 20 AAAGTAACGT TGACTTACCA ATACCTGGAT CCCCAcCAAT AAGTACTAAC GATCCGCTCA 6600  
 CAATACCTCC ACCTAATACA CGGTTGAATT CTGCTGAATC TGTTAAcACT CTCGGCGTTG 6660  
 TTTcATGtTT AATACTATTT AATTTTTGTA CTTTACCTGC TAATTCCTTG GTTTTAACTC 6720  
 25 CATGTTTAGG ATTGGCTGCT TTTTCAACAA TTTCTCCAT TTGATTCCAA GCGCCACAAT 6780  
 TAGGACATTT CCCCATCCAT TTAGGAGATT GATAACCACA AGCCATACAT TCAAAAATCA 6840  
 CTTTTTCTT GGCCArAATT GCACCTCCAC TTTCTT 6876

(2) INFORMATION FOR SEQ ID NO: 187:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1193 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

CAACTCAAAC AGCAGAACAA CGTCGTGAGT TGATTAATGG TGTATTTACT GACATTAATC 60  
 CCATACATTA AAAATATGAT GTACGTGTTA GCAGATAATA GACATATCTC ATTAATAGCT 120  
 45 GACGTATTCA AGGCGTTCCA AAGCTTATAT AACGGACACT ACAATCAAGA TTTTGCAACA 180  
 ATTGAGTCAA CATATGAATT GAGTCAAGAA GAGTTAGATA AGATTGTCAA ACTAGTAACT 240  
 CAACAAACGA AGTTATCTAA AGTTATTGTA GATACAAAAA TTAATCCAGA TTTAATTGGT 300  
 50 GGATTTAGAG TTAAAGTCGG CACAACtGTA TTAGATGGTA GTGTTAGAAA TGATCTTGTC 360  
 CAATTACAAA GAAAATTTAG AAGAGTTAAT TAATTATAAA GAGGAGTGAC ATAGATGGCC 420

ATGTCCGTAA CTGATGTAGG TACTGTATTA CAAATTGGTG ATGGTATTGC ATTAATTCAC 540  
 GGATTAAATG ACGTTATGGC TGGTGAGCTA GTAGAATTCC ATAACGGCGT ACTTGGTTTA 600  
 5 GCCCAAAACC TTGAAGAGTC AAACGTGGGT GTGGTTATTT TAGGACCATA CACAGGTATT 660  
 ACTGAAGGTG ACGAAGTTAA ACGTACTGGT CGTATCATGG AAGTACCAGT AGGTGAAGAA 720  
 10 CTAATCGGAA GAGTTGTAA TCCATTAGGA CAACCTATTG ATGGACAAGG ACCGATTAAAC 780  
 ACAACTAAAA CACGTCCaGT AGAGAAAAAA GCTACTGGTG TAATGGATCg TAAATCACTA 840  
 GATGAGCCAT TACAAACAGG TATCaAGCA ATTGATGCTT TAGTACCAAT TGGTAGAGGT 900  
 15 CAACGTGAGT TAATCATCGG TGACCGTCAA ACAGGTAAAA CAACAATTGC AATTGACACA 960  
 ATTTTGAACC AAAAAGATCA AGGTACGATT TGTATCTATG TTGCTATTGG TCAAAAAGAT 1020  
 TCAACAGTAA GAGCAAATGT TGAAAAGTTA AGACAAGCAG GCGCTTTAGA CTACACTATT 1080  
 20 GTTGTAGCAG CATCAGCTTC TGAACCTTCT CCATTATTAT ATATTGCACC ATATTCAGGT 1140  
 GTAACAATGG GTGAAGAATT CATGTTTAAAC GGTAACATG TTTTAATCGT TTA 1193

## (2) INFORMATION FOR SEQ ID NO: 188:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5549 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

35 TGCTAAGAAG TCAAAATAAA CTAActATna AACATCTAGT ACGATTATTA AAGTGACAGA 60  
 TnATAAAATT GAATTATtnA GAGAAGGAGA TATAAAGTTT GAAGAAATAA AAGAAAGACT 120  
 AGGTACAGGT ATTATTTATG AATAAGTTAA TACTTGGGAT TTATTTATAC CGAATTTTTT 180  
 40 CACGAGCATA CTTTTATTTA CCGTTTTTAT TAATTTACTT TTTGATTCAA GGTATTCCA 240  
 TAATACAATT AGAAATATTA ATGGCGTCTT ATGGCATTGC AGCATTTTTTA TTCTCTCTAT 300  
 ACAAGAGAGAA GTGTTTTTAAA ATTTGTAACT TAAAAGATTC TAATAAATTA GTTGTTAGTG 360  
 45 AAATATTCAA AATCATCGGT TTATTGTGTG TATTATATCA AAATCAATAT TTAATTTTAG 420  
 TAGTGGCACA AATATTATTA GGGTTAAGTT ACTCAATGAT GGCGGGTGTT GATACCGCAA 480  
 TAATTAAGAG AAATATAACA AATGAGAAAT ACGTACAAAA TAAGTCAAAT AGCTATATGT 540  
 50 TCCTATCATT ATTAATTTCa GGGATTATAG GTAGTTATCT TTATGGAATA AATATTAAAT 600  
 GGCCTATAAT AATGACTGGT ATATTTTCAA TTCTAACAAT TATAATTATT CGATGCACAT 660

55

	TACCAGAAGA GAAGTTTGG ATATTGCATT ATTCTTTTTT AAGAGCGTTA ATATTAGGAT	780
	TTTTTATAGG ATTTATTCCA ATTAATATAT ATAATGATTT AAAACTGAAT AATTTACAAT	840
5	TTATTTCACT ATTAACCTGT TACACAGTTA TGGGTTTTGT ATCTTCACGT TATTTAACTA	900
	AATACTTGAA TTATAAGTTT GTGTCAGAAA TTTGTTTAGT AATATTTTTA ATAATATATA	960
10	CATATCAAAG TTTCATAGCA GTTACTATTT CTATGATATT TTTAGGTATT TCTTCAGGGT	1020
	TAACTCGTCC ACAAACCTATA AATAAACTTT CTAGCAGTAG TAACTTAAGA GTGATGCTTA	1080
	ATTATGCAGA AACGTTATAT TTTATTTTTA ATATCGCATT TTTACTTATG GGTGGTTACT	1140
15	TATATACAAT AGGAACTATT CAATACTTAA TATTATTTAT TTCGTTATTA ATTTTTATAT	1200
	ATTTAATAAT AATATTTTAT TTTACAAGGA GAGAGCAACA TGAAAATAAA AACTGAATTT	1260
	AAAGGGAACA ATATACCATA TGAATACGCA GCAGGTGCAG ATGTGAGTGA TTCTATTAAC	1320
20	GGGAATCCAA TTAAGTCATT TCCATTTGAA GTAATTGAAT TACCGGAAGG gACTAAATAT	1380
	CTTGCTTGGT CTTTAATTGA CTATGATGCA ATTCCTGTAT GTGGCTTTGC TTGGATTTCAT	1440
	TGGAGTGTAG CTAATGTAAG TGTTAGTGGC AATTCATTTT CTATAAAAGC AGATTATCA	1500
25	AGAACAAAGG GCGACTATGT ACAAGGTAAA AATAGCTTTA CTAGTGGGTT GTTGGCTGAA	1560
	GATTTTTTCAG AAATAGAAAA TCACTATGTA GGACCTACAC CACCTGATCA AGATCATCAA	1620
30	TATGAATTAA CAGTTTATGC GTTAGATCAT TCTTTAAATT TGAAGAATGG GTTCTACTTG	1680
	AATGAATTTT TAAAAGAAGT AAATCAACAT AAAATTGATC AAACAAGTAT TAACCTTATA	1740
	GGAAGAAAAA TTTAATACTA AATATCTCAT CAATATAAAA TTGTTCAATT AAAAGTACAA	1800
35	AGAAACAAAG GTTTTAATTT ATATATTAGG TACGGCGTTC GCTATAATGC AAAGAAGTAA	1860
	TTAAATTTAA GAAATGTAAA CTTAGTTATT GTAATGTGAA TTTATTTGAA AAAATAGAAA	1920
	GTATTAACAA TTATAGCTTT TACATTAATT AAAATTTATT TTTAAAAACA AGTAAACAAT	1980
40	TTACATACTT ATAATTTTTG AAAATTTTCA ATTTGTGTTA TATTGATTTT GTAAGATACT	2040
	TTAACTCACA AAGGAGAGAG AGTATATGAA ATTAAATCA TTTATAACTG TAACTTTGGC	2100
	ACTGGGCATG ATCGCAACGA CTGGCGCTAC TGTGGCAGGT AATGAGGTAT CTGCAGCAGA	2160
45	AAAGGACAAA CTACCGGCAA CTCAAAAAGC TAAAGAAATG CAAAATGTTC CATATACAAT	2220
	TGCAGTAGAT GGCATTATGG CTTTCAATCA ATCTTACTTA AATTTACCAA AAGATAGCCA	2280
50	ATTATCATAT TTAGATTTAG GAAATAAAGT TAAAGCTTTG TTATATGATG AACGCGGTGT	2340
	AACACCTGAG AAGATTCGAA ATGCAAAATC TGCCGTTTAC ACGATTACTT GGAAAGATGG	2400
55	TAGTAAAAAA GAAGTGGATC TTAAGAAAGA TAGCTACACA GCAAACCTGT TTGATTCAAA	2460

	CAACATGAAG CATTTAATTT TACAGTGATG ATTATAAAAT AATTGCCTTG ATACAAAGAT	2580
	TACTCGTAAA TGACATCTTT GTATTAAGGC TTTTCTAAA TTTAAAAGTG ATGGGTTAGA	2640
5	GGTCATTGAG CTTTAAAATA TTCAAATAC AAAACATTAA TGGCCAAAAA TAAAGCCGC	2700
	CTTTATCTGG GCAGCTTCAA TAATAAGAAA GACATATTTC ATTTTATACT AAATAGTTAT	2760
	TGTGATGAAT CTTTCGGCGG TTTAATTACT GCAGCAAAAA TTGCTGTGAA AATCGTGAAC	2820
10	AATACTGCCA TGATAATTGG ATTCACTACA TTTAAGCTGT CTCCACCTAC TAGGCTATTA	2880
	AGTACAAAGT TAACCATTG CATTAATAAT AATGCCCAAA AGAATGTTAC GAGGTGTTTC	2940
15	ATGTCATTCT ACCTCCACTT TAATTATATA TATTTTATTT TAAGTGAAAG TTAGAAATTT	3000
	GTATAGTAAC ATCTCATATA TTTTGACCAT ATTATACAGT TTAAATAAAT GATTTTATCT	3060
	GAATGGCTAT TCTAAATTAA GCGCATTAAA ACCAATTTCA TACTGAAATT TGACGATAAT	3120
20	AAAGCATTAA AATTTTATTA ACTAGTCAAT ATTCTACCT CTGACTTGAG TTTAAAAAGT	3180
	AATCTATGTT AAATTAATAC CTGGTATTAA AAATTTTATT AAGAAGGTGT TCAACTATGA	3240
	ACGTGGGTAT TAAAGGTTTT GGTGCATATG CGCCAGAAAA GATTATTGAC AATGCCTATT	3300
25	TTGAGCAATT TTTAGATACA TCTGATGAAT GGATTTCTAA GATGACTGGA ATTAAAGAAA	3360
	GACATTGGGC AGATGATGAT CAAGATACTT CAGATTTAGC ATATGAAGCA AGTTTAAAAG	3420
	CAATCGCTGA CGCTGGTATT CAGCCCGAAG ATATAGATAT GATAATTGTT GCCACAGCAa	3480
30	CTGGaGATAT GCCATTTCCA ACTGTGCAAA ATATGTTGCA AGAACGTTTA GGGACGGGCA	3540
	AAGTTGCCTC TATGGATCAA CTGTCAGCAT GTTCTGGATT TATGTATTCA ATGATTACAG	3600
35	CTAAACAATA TGTTCAATCT GGAGATTATC ATAACATTTT AGTTGTCGGT GCAGATAAAT	3660
	TATCTAAAT AACAGATTTA ACTGACCGTT CTA CTGTCAGT TCTATTTGGA GATGGTGCAG	3720
	GTGCGGTTAT CATCGGTGAA GTTTCAGATG GCAGAGGTAT TATAAGTTAT GAAATGGGTT	3780
40	CTGATGGCAC AGGTGGTAAA CATTTATATT TAGATAAAGA TACTGGTAAA CTGAAAATGA	3840
	ATGGTCGAGA AGTATTTAAA TTTGCTGTTA GAATTATGGG TGATGCATCA ACACGTGTAG	3900
	TTGAAAAAGC GAATTTAACA TCAGATGATA TAGATTTATT TATTCCTCAT CAAGCTAATA	3960
45	TTAGAATTAT GGAATCAGCT AGAGAACGCT TAGGTATTTT AAAAGACAAA ATGAGTGTTT	4020
	CTGTAAATAA ATATGGAAAT ACTTCAGCTG CGTCAATACC TTTAAGTATC GATCAAGAAT	4080
50	TAAAAATGG TAAAAATCAA GATGATGATA CAATTGTTCT TGTCGGATTG GGTGGCGGCC	4140
	TAACTTGGGG CGCAATGACA ATAAATGGG GAAAATAGGA GGATAACGAA TGAGTCAAAA	4200
	TAAAAGAGTA GTTATTACAG GTATGGGAGC CCTTCTCCA ATCGGTAATG ATGTCAAAAC	4260
55		

TGAACCTTAT AGCGTTCCT TAGCAGGAGA ACTTAAAAAC TTTAATATTG AAGATCATAT 4380  
 CGACAAAAAA GAAGCGCGTC GTATGGATAG ATTTACTCAA TATGCAATTG TAGCAGCTAG 4440  
 5 AGAGGCTGTT AAAGATGCGC AATTAGATAT CAATGAAAAT ACTGCAGATC GAATCGGTGT 4500  
 ATGGATTGGT TCTGGTATCG GTGGTATGGA AACATTGAA ATTGCACATA AACAAATTAAT 4560  
 GGATAAAGGC CCAAGACGTG TGAGTCCATT TTTCTGTACCA ATGTTAATTC CTGATATGGC 4620  
 10 AACTGGGCAA GTATCAATTG ACTTAGGTGC AAAAGGACCA AATGGTGCAA CAGTTACAGC 4680  
 ATGTGCAACA GGTACAAATT CAATCGGAGA AGCATTTAAA ATTGTGCAAC GCGGTGATGC 4740  
 15 AGATGCAATG ATTACTGGTG GTACAGAAGC ACCAATTACT CATATGGCAA TTGCTGGTTT 4800  
 CAGTGCAAGT CGAGCGCTTT CTACAAATGA TGACATTGAA ACAGCATGTC GTCCATTCCA 4860  
 AGAAGGTAGA GATGGTTTTG TTATGGGTGA AGGTGCTGGT ATTTTAGTAA TTGAATCTTT 4920  
 20 AGAATCAGCA CAAGCTCGAG GTGCCAATAT TTATGCTGAG ATAGTTGGCT ATGGTACTAC 4980  
 AGGTGATGCT TATCATATTA CAGCGCCAGC TCCAGAAGGT GAAGGTGGTT CTAGAGCAAT 5040  
 GCAAGCAGCT ATGGATGATG CTGGTATTGA ACCTAAGAT GTACAATACT TAAATGCCCA 5100  
 25 TGGTACAAGT ACTCCTGTTG GTGACTTAAA TGAAGTTAAA GCTATTAAAA ATACATTGTTG 5160  
 TGAAGCAGCT AAACACTTAA AAGTTAGCTC AACAAAATCA ATGACTGGTC ACTTACTTGG 5220  
 TGCAACAGGT GGAATTGAAG CAATCTTCTC AGCGCTTTCA ATTAAAGACT CTAAAGTCGC 5280  
 30 ACCGACAATT CATGCGGTAA CACCAGATCC AGAATGTGAT TTGGATATTG TTCCAAATGA 5340  
 AGCGCAAGAC CTTGATATTA CTTATGCAAT GAGTAATAGC TTAGGATTCTG GTGGACATAA 5400  
 35 CGCAGTATTA GTATTCAAGA AATTTGAAGC ATAACATAA AATCTTCAG TAACGTTGTT 5460  
 TTAGTTACTG AAGATTTTTT CaGTTTCTTT ATACTAAGAT GAGCGACaCa CAATCGTCAT 5520  
 AATAAAATAT GAATATTTAT TAATAATAA 5549

40 (2) INFORMATION FOR SEQ ID NO: 189:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4832 base pairs  
 (B) TYPE: nucleic acid  
 45 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

AGATTATAGT AAGATTGATA GTTTGGCGAC TGaAGCgCGa GaAAAATTAT CAGaAGTAAA 60  
 mCCTTTAAAT ATTGCACAAG CTTCTAGAAT ATCAGGGGTA AATCCAGCAG ACATATCTAT 120

55

	TGGTTAGCAG AACAAATAAA AGAACATAAT ATTCAATTAA CTGAGACTCA AAAACAACAG	240
	TTTCAAACAT ATTATCGTTT ACTTGTGTAA TGGAAATGAAA AGATGAATTT GACAAGTATT	300
5	ACAGATGAAC ACGATGTATA TTTGAAACAT TTTTATGATT CCATTGCACC TAGTTTTTAT	360
	TTTGATTTTA ATCAGCCTAT AAGTATATGT GATGTAGGCG CTGGAGCTGG TTTTCCAAGT	420
10	ATTCCGTTAA AAATAATGTT TCCGCAGTTA AAAGTGACGA TTGTTGATTC ATTAAATAAG	480
	CGTATTCAAT TTTTAAACCA TTTAGCGTCA GAATTACAAT TACAGGATGT CAGCTTTATA	540
	CACGATAGAG CAGAAACATT TGGTAAGGGT GTCTACAGGG AGTCTTATGA TGTGTACT	600
15	GCAAGAGCag TAGCTAGATT ATCCGTGTTA AGTGAATTGT GTTTACCGCT AGTTAAAAAA	660
	GGTGGACAGT TTGTTGCATT AAAATCTTCA AAAGGTGAAG AAGAATTAGA AGAAGCAAAA	720
	TTTGCAATTA GTGTGTTAGG TGGTAATGTT ACAGAAACAC ATACCTTTGA ATTGCCAGAA	780
20	GATGCTGGAG AGCGCCAGAT GTTCATTATT GATAAAAAAA GACAGACGCC GAAAAAGTAT	840
	CCAAGAAAAC CAGGGACGCC TAATAAGACT CCTTTACTTG AAAAATAATG CATAATCCTT	900
	TACAATAAC ATAAAAGGAG CGAATGGATA ATGAAAAAC CTTTTTCAAA ATTATTTGGT	960
25	TTGAAAAACA AAGATGACAT CATTGGACAT ATTGAAGAAG ATCGCAATAG TAATGTTGAA	1020
	TCCATTCAAA TTGAACGTAT CGTTCCCAAC CGTTATCAAC CAAGACAGGT GTTTGAACCA	1080
30	AATAAAATTA AAGAACTTGC TGAATCAATA CATGAACATG GTTTACTACA ACCTATTGTT	1140
	GTAAGACCGA TTGAAGAAGA TATGTTTGAA ATTATTGCTG GAGAGCGCCG ATTTAGAGCA	1200
	ATACAATCAC TAAATTTACC TCAAGCAGAC GTTATTATTC GTGATATGGA TGATGAAGAG	1260
35	ACGGCTGTTG TTGCATTAAT TGAGAATATT CAAAGAGAAA ATTTGTCTGT TGTGAAGAA	1320
	GCGGAAGCCT ATAAGAAATT ATTGAAATT GGTGATACAA CGCAAAGTGA ATTGGCAAAA	1380
	AGTTTAGGTA AAAGTCAAAG CTTTATTGCA AATAAGTTGC GTTTATTGAA GTTGGCGCCG	1440
40	AAAGTACTAC TTCGCTTAAG AGAAGGTAAA ATTACTGAAC GTCATGCGAG AgcGGtATTA	1500
	TCATTGTCTG ATAGCGAACA AGAAGCGTTG ATTGAGCAAG TCATTGCACA AAAGCTAAAT	1560
	GTGAACAGAc TGAAGATAGA GTACGCCAAA AAACGGGGCC CGAAAAAGTC AAAGCACAAA	1620
45	ACCTTCGCTT TGCACAAGAT GTCACTCAAG CACGAGATGA GGTAGGCAAA AGTATCCAAG	1680
	CGATTCAACA AACAGGATTA CATGTTGAGC ATAAAGACAA AGATCATGAA GATTATTATG	1740
50	AAATAAAAAT TCGAATATAT AAACGTTaGT AGTAGGATGT CGTATACATG ATGACTAACA	1800
	CATAAAAGAC AAAGCTAAGA TCATAACAGC TTTGTCTTTT TTTTTTGTTF TACGTGAAAC	1860
55	ATAAAAATTT ATATTTATAT GTTGATCAGG CTGGTACATA AATCAATGTT CTATGCTCTA	1920



	TTCTAGTCAA CCTTGCTGGG GTGGGACGAC GAAATAAATT TTGCGAAAAT ATCATTTCTG	2040
	TCCCACTCCC TAATTTGAGC TGGATATACT TTCATTTGAA CCCTTTATTG CTAGTTTATG	2100
5	AAAGTATCAT GAAAGCTTTA TGAACATCGC TTGAGTTGCC TTTACAGTAG AAAATTTAAG	2160
	TTTTACACTT TGTGTGAATG ATACGTTTTG TATTGAATTA ATTATAGAAA GGTACGTTGA	2220
	AGATGTTTTC AATTGGAAGT GCAATTCCTC ATTTTGTTCAT TGGTGGTATC GCTGTTGCAT	2280
10	TAGCTTCAAT TATTGCTGAT AAGGTAGGTG GTAAGTTAGG AGGTATTATA GCTACTATGC	2340
	CGGCAGTCTT TCTTGGCGCT ATTATCGCAT TAGCTTTAGA TCATCGTGGT ACGCAATTAG	2400
15	TGGAGATGTC GATGAATCTT AGTACTGGAG CAATTGTCGG TATTCTGTCT TGTATATTAA	2460
	CTGTATTTTT GACATCTCTC TACATTAAGC ATAAAGGTTA TCGGAAAGGC GCAATATTCA	2520
	CAGTTGTTTG TTGGTTTGTC ATTTCCCTCG CAATATTCAG TATTAGACAT TTATAGTTTG	2580
20	GAAAATGCGT GATAATTAGT TGTATTCAGT TATTAAGTAA TAAATTATTG GAGGCAGAAC	2640
	ATCATGAAAT TAACATTAAT GAAATTTTTT GTGGGGGGAT TTGCAGTATT ATTAAGTTAT	2700
	ATTGTATCTG TAACACTACC TTGGAAAGAA TTTGGCGGTA TATTTGCaAC GTTTCGGGCA	2760
25	GTATTTTTAG TGTCTATGTT TATTACAGGT ATGCAATATG GTGATAAAGT CGCTGTGCAT	2820
	GTAAGTCGTG GCGCAGTGTT TGGTATGACA GGGGTATTAG TTGTATTTT AGTTACATGG	2880
	ATGATGTTAC ATATGACGCA CATGTGGTTG ATTAGCATTG TTGTTGGTTT CCTAAGCTGG	2940
30	TTCATCAGTG CAGTATGTAT TTTTGAAGCG GTAGAATTTA TAGCACAAAA AAGATTAGAA	3000
	AAGCATAGTT GGAAAGCTGG AAAATCGAAT AGTAAATAGT GTGAACGTAA TCTCTTAACT	3060
35	AGGACTAACT TTGCAAGCAT TGAATAGCAT GGAAAAGTTG CATCATTAAT AAGTGAAATT	3120
	CAAGTTGGCA TTGAGAAAAT TACAAGCGCG TAATCATACa GGTCTGTCTT AAGGGAGTCT	3180
	TCGAACCCCG ATGTTGTCGT ATGTCAAAAC ATTTAGTCAA TCATAAAGGT GACTTGATTT	3240
40	AACTTTATCT GATAGTCTGA TTGTAATGAT TGTACTAATT GACTGGAGGC GTATGTAATT	3300
	GAATCTGAGT AAACAAATTA AAAAGTATAG GGAACGAGAT GGTATTTCAC AAGAATATCT	3360
	TGCTGAAAAG TTATATGTAT CTAGGCAGAG TATTTCTAAT TGGGAAAATG ACAAAGCTT	3420
45	ACCAGACATA CATAACTTAT TAATGAYGTG TGAATTGTTT AATGTAACCT TAGATGATTT	3480
	AGTAAAGGG ACCATTCCAT TTGTACCTGA TATTAAAGCG CAACGAAGTC TTAACCTATG	3540
	GACATATGTG ATGCTTATTT TCATGACATT AGCTGCAATT TTAATGGGAC CTTTAGTTGT	3600
50	TTATTGGAAT TGGACTTGGG GTGTAACGGT GGCAATCATT TTGGGAATAG GTTTTTATGC	3660
	ATCTATGAAA ATAGAAGATT TAAAAAAGT GCATAAAATG GACAACTACG ATCGAATTGT	3720

55

5 GACAAATGCG CTTTCTATTA TATCAGTAAT TGGTATACTC AGCCTCATAA TTTTCCTTAG 3840  
 TGTGTATTTG GCAAATAAGT TTTTATAAAT CATCGTGGTA TCGTCTCATA TTATTTATAT 3900  
 10 TATCCAAAAT AGCATAAAAA AATACCAACA AGATTTAGAA CCTTGTTGGT AATCAAAGCG 3960  
 aTTCATTTAT AATGAGTCGT TTTATGTTGT AAGATTAAAC AGTTTGTACG TTAAGTCTT 4020  
 GGTCTCCACG TTGACCTTCA GTGATTTTCA AAGTAACTTT TTGACCTTCT TCTAAAGTTT 4080  
 15 TGTAGCCATC GCTAGCGATA CCTGAGAAAT GTACGAATAC GTCTCCGCCA TTTTCTTGTT 4140  
 CGATGAAACC AAAACCTTTT TCTGCTTTAA ACCATTTWAC TGTACCGTTA TTCATATWGA 4200  
 AwACCTCCGT gTGCTTTTGC ACTTAATATT TGTAACAAAT TCATAACTAA AAAAGAGGAT 4260  
 ATTCTAAACA AATACACTAC AATTTAATTC ACGAGCTTTT ATTACGTAAG ACCAACTATA 4320  
 CGCTCATATT GGCATAATGT ACAGTGT TTTT TGAATAA ATTAATAAAG ATTTTAAAA 4380  
 20 ACCTTAGAAA CGTTGATTTA AAGGGGTTTA TAAAAATwAw AAAATTGTAG TCTTTTATGG 4440  
 TGTGTGCTAG TTTTCAAAGT GACATATCGT TTAACATGA TGATTTTATA AGCAATCCAT 4500  
 AAAAAACAAG CAGCGATAAA CGCTACTTGT TGATATTAA ATCTGACTTG AAAGGTCATA 4560  
 25 GCAATGTTCT ATACCGATGG AATGTGCTTA CTTGCCTTTT TCTTCACGAC GTTTTAAATA 4620  
 ATAAGAGCCA CCTAATAAAC CAGCTGGAAT GCCTATCATT GGTGTTGTGA ATGAGCTTAA 4680  
 TACAATAACA AGTATTGTTA AAGCAATGAC GTTATACCAA GTTACAGTCA AATTTTTCAA 4740  
 30 ATCCTCATAT GATTGTTTTA CTAATCTCT AAATTCATG ATTCAATCTC TCCTTTTTTA 4800  
 TAAATCTTTA GATTGTCAA TTAAGCTGGA CA 4832

35 (2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5727 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

45 CAAAGCTGTT CAAAAGGCTT ATAATTTAAA TTTAGATAAC ATACGTACAA TGGAACCTAA 60  
 GTTGAGATAT CAAGCGATCA ATAAAGGTAA TATTAATTTA ATAGATGCAT ATTCAACTGA 120  
 CGCTGAATTA AAACAATATG ATATGGTTGT GTTAAAAGAT GATAAGCACG TATTTCCACC 180  
 50 ATATCAAGGA GCACCATTAT TTAAAGAAAG CTTTTTAAAG AAACATCCAG AAATTAAGAA 240  
 ACCGTAAAC AACTAGAAA ACAAATATC TGATGAAGAT ATGCAAATGA TGAAGTATAA 300

55

	GTTAATCAAA TAACGACCAA CGCCACATAA GATGCGTAAC ACCAAATTAT ATCTTATGTG	420
	GCGTTGTTAT ATTTAAATCT ATAATTATGT TCAATTTAAA CATGCAATAA TGATTAAAAA	480
5	ATATGACATG TTAAACACAA TGTAAGCTAT TATGATGTGA AAATAGTAGC ATTGCATTTT	540
	AGAAACATAG AGCGATATAA TGAATATAAG TTTTTTGAAA TTTCAGTTAA TTCTAAGGAG	600
	GTTGTTTTTA TTATGAAAGA ACAACTTAAT CAACTATCAG CATATCAGCC TGGTTTATCT	660
10	CCAAGGgCAT TGAAAGAAAA GTATGGCATT GAAGGAGATT TATATAAACT TGCATCAAAT	720
	GAAAATTTGT ATGGACCATC GCCTAAAGTT AAAGAAGCGA TATCAGCACA CTTAGATGAG	780
15	TTATATTATT ATCCTGAAAC AGGATCACCG ACATTAAAAG CGGCGATTAG TAAACATTTA	840
	AATGTAGATC AATCACGCAT TTTATTTGGT GCGGGATTAG ATGAAGTTAT ATTAATGATT	900
	TCTAGAGCTG TATTAACGCC AGGGGATACT ATTGTTACAA GTGAAGCGAC ATTCGGTCAA	960
20	TATTATCACA ATGCGATTGT TGAATCAGCT AATGTGATAC AAGTACCTTT AAAAGATGGT	1020
	GGCTTCGATT TAGAAGGTAT TTTAAAAGAA GTTAATGAAG ATACGTCATT GGTATGGTTA	1080
	TGTAATCCAA ATAATCCTAC AGGTACATAT TTAATCATG AGAGCTTAGA TTCGTTTTTA	1140
25	TCTCAAGTAC CTCCACATGT ACCAGTAATT ATAGATGAAG CTTATTTTGA ATTTGTGACA	1200
	GCAGAGGACT ACCCGGATAC ACTTGCTTTG CAACAAAAAT ATGACAATGC TTTCTTATTA	1260
	CGTACATTTT CAAAGGCGTA TGGATTAGCG GGTTTACGTG TAGGATATGT GGTAGCAAGT	1320
30	GAACATGCGA TTGAAAAATG GAACATCATT AGACCACCAT TTAATGTGAC ACGTATATCT	1380
	GAATACGCAG CAGTTGCAGC ACTTGAAGAT CAACAATATT TAAAAGAGGT AACACATAAA	1440
35	AATAGTGTTG AACGCGAAAG ATTTTATCAA TTACCTCAA GTGAGTATTT CTTGCCAAGT	1500
	CAAACGAATT TTATATTTGT AAAAACmaAG CGGGTAAATG AACTTTATGA AGCACTTTTA	1560
	AATGTAGGGT GTATTACGCG ACCATTTCCA ACTGGTGTTA GAATTACAAT TGGTTTTAAA	1620
40	GAACAAAATG ATAAAATGTT AGAAGTTTTA TCAAACTTTA AATACGAATA GTAAGTGGGG	1680
	AGTGGGACAG AAATGATATT TTCGCAAAT TTATTTCTcC GTCCCACCCC AACTTGcATT	1740
	GTCTGTAGAA ATTGGGAATC CAATTTCTcC TTGTTGGGGC CCCGCCGGCA AGGTTGACTA	1800
45	GAATTGAAAA AAGCTTGTTA CAAGCGCATT TTCGTTcAGT CAACTACTGC CAATATAACT	1860
	TTGTAGAGCA TTGAACATTG ATTTATGTCT CAAGCTCAAT GCAGTGTGAA TGATGAGGTG	1920
	AGAGTATTCA GTGTAAAAAG CAACAATAGA TGATATTGTT TTGTATCAAT TGCTTTTTTG	1980
50	CTATACTGAA TCAATACTGA TATTTTCAGG AGAAGATTAA AATGACCCGT AAATCAATCG	2040
	CGATTGATAT GGATGAAGTA TTGGCAGATA CATTAGGAGA AATCATTGAT GCTGTCAATT	2100
55		

	TTCCTGAACA	TGATGGATTA	ATTACAGAAG	TATTGAGAGA	ACCAGGCTTC	TTCAGACATC	2220
	TTAAAGTGAT	GCCGTATGCA	CAAGAAGTTG	TGAAAAAATT	AACTGAACAT	TATGATGTAT	2280
5	ATATTGCTAC	AGCAGCAATG	GATGTACCAA	CATCATTTAG	TGATAAATAT	GAATGGTTAC	2340
	TAGAGTICTT	TCCATTTTTA	GATCCTCAGC	ATTTTGTTTT	TTGTGGTAGA	AAAAACATCG	2400
	TTAAAGCTGA	TTATTTAATA	GATGACAATC	CTAGACAGCT	TGAAATTTTT	ACTGGTACAC	2460
10	CGATTATGTT	TACAGCAGTG	CATAATATTA	ATGATGATCG	ATTTGAACGC	GTAAATAGCT	2520
	GGAAAGATGT	AGAACAGTAT	TTTTTAGATA	ATATTGAGAA	ATAAAATATA	TCACTTGAAA	2580
15	AATTTTCATGT	AGAAAAGATG	ATGGATAGGC	TATAAAGTAA	TTGTGACTGA	GATGAACTTT	2640
	TATGTCTTAG	ACACTACAAC	ACTATATTGG	CAGTAGTTGA	CTGCGGGGCC	CCAACATAGA	2700
	GAAATTGGAT	TCCCAATTTT	TACAGACAAT	GCAAGTTGGG	GTGGsCCCCA	ACATAAGAA	2760
20	ATACTTTTTT	TTTAGAAAAT	AGTATTTCTT	ATGCATGAGT	GTAATCATG	CATTCATATT	2820
	TTTAAGTACA	CATTAGCTGT	GACTAATGAT	AAAGAATCGC	TACATAATCA	ATCATTAGTC	2880
	GTTCTTTATC	ATTTCCGTCC	CGCTCTCAAT	AAATGTTAGT	CTATCTTATT	ATTATAAATC	2940
25	GGATGAATGT	GTTAATCTAT	GGCAGATTAC	ACGTCATCCG	ATTTTTTATA	GAATTTGAAA	3000
	AAGACGCATA	AACCACTATG	ATTTAAAATA	CAACATCAAT	CATTTTAGTG	gCATGCGCCA	3060
	AAATTATATG	TCTGTTTTTG	AAACAGGGTA	ATAGCTTAAA	GCTAATAAAA	ACGAATATAA	3120
30	GGTGCCTTGA	ATCTTATGAT	TACACTCCAA	ACCTAATATA	ATATCGGGTT	AAGATCATTC	3180
	CGGATGCTTA	CAAATCATTG	ACAGTAAGTA	ACTGAATGGC	ATTTGGTATA	ACCTCAATAT	3240
35	CAATAGGTGT	TTCTAATGAA	ATTTCGCCAT	CAATATCAAC	TTTCATTGCT	GGATCTGTTG	3300
	TAAGTGAAAT	CTTTTTACCA	GGTATATGCT	CAATACCTTG	AGTAATTTCA	TTCCaATTCA	3360
	TGCTATCACG	CTTTTTAAAA	ATATCATTTA	AAATACTGAA	ACTTTGTTCA	TTAAAAATGA	3420
40	AAGTGTTTCA	TTCACCATCT	TGAGGAGACA	AATCAGTCaA	TGGTATACGA	CTACCACCAA	3480
	TGAATGGACC	ATTTGCTGTT	AGTATCATGG	TCGTTTCGCC	AGAATATGTC	TTATCATCTA	3540
	TTGATAATTG	ATAATTAAAT	TGTGTTGGAT	TTAGCAGTGT	TTTGACAGTT	GATCCAATAT	3600
45	AACTCAATTT	ACCAAATATA	TCTTTTGAAC	CATCTTGTAC	GTTTTCAGCG	TTTTGAACAA	3660
	TGAGACCTAA	GCCAACAAAG	TTGAGTGCAT	ATTGATTATT	TATTTTAATT	ACATCGTATG	3720
	TACCAACTTG	TGCAGAAATC	ATTTGTTTAC	TAGCTTGTTC	ATGATTAGGT	GCTATATTTA	3780
50	GCGTTTTTGT	AAAATCATTA	AAAGTACCGC	CTGGTAAAAT	GCCAATAGGG	AGTTGAAGGT	3840
	CATGTGTCAT	AACACCGTTT	ATAAGTTCGT	TAACCGTGCC	ATCACCGCCA	AGAATAAATA	3900

55

	CACCTTCGTT	TTCACCTCAAT	TGAATAGAAA	GATGCTTACA	AATTGAACTT	AATGCTGTTG	4020
	TAACTTCCCC	AATACCTTGA	TTAATATTTT	TTAATCCACT	GTGTTTCATGG	TAAAAGAGGA	4080
5	CACCATGTGT	ATATTTATTT	TCCATAGTTT	AGCCTACTTT	CTAAAAATTG	GTTCAATTAAA	4140
	TATATATACC	CACCTTTAAT	TGTTAATACC	AAAAATATGT	TTTTAAATAG	AGAAAATGGT	4200
	AATAAATGAA	ATTGATTTCT	ATAGAGTGGG	ACGAGAAAAT	ATAGTTATAG	CTGCTATATA	4260
10	TGAGCATATT	AAGTTTTTAT	TTATACTGAT	ATCTTGAATT	TAATTAATAG	AAACCTATAA	4320
	AAAAACAGTA	AGCCATTTAA	ATGACTTACT	GTTTTTTGAA	TTAGGCCAAC	AATATTAACG	4380
15	TATACCTTTC	ATCGCTTTGA	TGATTAAAGG	TGAGAATGCT	AATACAATTG	TTGTAACAAAT	4440
	AATTGCAACA	ACACCTAGGA	AAATAAAGTA	ATTGTTTGA	CCTAGTGGTT	CTATTAACCTT	4500
	AACTAAAGTA	CCATTGATTG	CTTGTGCAGA	AGCGTTAGTT	AAGTACCAA	TACTCATCAT	4560
20	TTGGGCATTA	AATGCTTTAG	GTGCTAACTT	AACAGCAGCA	CTATTACCCG	TTGGTGATAA	4620
	GCATAGCTCA	CCGATAACAC	AAATAATGTA	CGATAAAATA	ACCCAGTTAA	CTGAAAAGTT	4680
	TGATGAACCT	GATGCATAAC	CTACAATACC	AATTAGTATG	TATGACGCAC	CTGCTAAGAA	4740
25	CGTACCAATT	GCAAATTTTA	CTGGCAGGCT	AGGTTGTTTA	GTTCCAAGCT	TTTGCCATAA	4800
	AAGTGAAATA	ATTGGAGCTA	GTAATAAAAT	AAATAATGGG	TTAATTGATT	GGAAGATCGC	4860
	TTCACCAAAG	TTTGTTTTCC	AACCAAATAA	GTTTAATTTT	ATATCTGAAT	GTTCAATTCC	4920
30	ATATATGTTT	AATACATTAG	ACCTTGTTTC	TTGAATAGCC	CAGAACACCA	TTCCAAGAAT	4980
	AAATAATGGA	ATAAATGCTT	TAACACGAGA	ACGTTCAAGT	TCAGTGACAT	CTTTACTTCT	5040
	AAATAATTAAA	GTGAAGTAAA	TGAATGGTAA	TGCAATACCT	AATACTAAAA	CAGTATTACT	5100
35	AACTAAGTTA	AATGATAATG	AGTTAGTTAA	TGCACCAATA	ACGATAATTA	ATACAATTGC	5160
	TAAACAACA	CTTCCGATAA	TAAGACCATA	CTTTTCTTTT	TCAGCTGGTG	TCAATGGGTT	5220
40	AGTAGGTTTC	ATACCAACGC	TACCTAAGTT	TTTGCGGTTG	AAAAGTACAT	ACCATACTAA	5280
	ACCTAATGCC	ATACCAACTG	CTGCAATCAA	GAATCCGCCG	TGGAAGTTTT	TAACATTAAAC	5340
	AAAGTGTTGC	AAAATAATAG	GTGATAATAA	TGCACCCATA	TTAACTGACA	TATAGAAAAT	5400
45	AACAAAACCT	GCATCCATAC	GTCTATCATT	TTCAGGATAT	AAACGGCCAA	CGATATTTGA	5460
	AATGTTTGGC	TTCAATTAAAC	CTGAACCAAT	AATGATGAAG	AACATTGATG	TGAATAAGCC	5520
	GATTAATGCA	AATGGTAAGC	TTAAACAAAT	ATGTCCGATA	ATAATAAAGA	CTGCACCTAA	5580
50	TAAAGTAGCG	CCTCTAGTGC	CTGTAATTCT	GTCAGCAATC	CATCCGCCTG	GTATTGATGT	5640
	CATATAGATT	AATGAACCAT	AACTGACAT	AATTGACATA	GCTGTTGTTT	TATCAATTCC	5700
55							

## (2) INFORMATION FOR SEQ ID NO: 191:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14078 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

	TGGA	CTATTA	ACGGCGa	AGA	AGATTTA	ACG	AAATACT	TAC	AAACCAAT	GT	TGATGGT	TATT	60
	ATCACAG	ATG	ACCCAGC	ATT	AGCTGAT	CAG	ATTAAAGA	AAG	AAAAGAA	AGA	CGAAACAT	AC	120
	TTCGAT	CGTT	CTATAAGA	AAT	TTGTTT	TGAA	TAATATAA	AC	AAAGACCT	CT	AAAGTTAT	CA	180
	AGATGATA	ACC	TTCAGAGG	TTC	TTTTAAT	GT	TGCCATCT	TAT	GGGATAGG	CA	ATCGTTTC	AT	240
	TCGTTT	TATAT	TCATATG	ACA	AGTATTT	TGTA	TGGCAATT	TG	GCGTCACA	AAA	CACTTAC	ATG	300
	ATTTATT	GGT	GAATTATT	A	TTGTTTT	TGTG	AATGCAA	AGG	GTTAGAA	ATT	GAATTGT	AAA	360
	TACTTT	CTAA	TCTTTG	TTTC	GCTTTAG	TCA	TTTGATCC	AA	ATTTTTAG	TG	CGTATAG	CGG	420
	ATTTTG	CAAT	ATAGTG	CGCA	cTAAAA	TATC	GCGTTTTT	TGA	AACGCAT	CTA	AATTTAG	GTA	480
	CGATAAT	TTA	TTTAAGT	CAG	TGTTTG	CTAT	TAATTCAT	GT	AATTGAT	CTA	CAAGCGCT	TG	540
	ATGTTG	AAC	GTATGT	GATG	TAGTTTC	CAGA	TTTGCTTG	CT	AATTTA	AATAC	CAGTCGT	ATC	600
	AAGGAG	CGCC	GCTTTA	ATAC	CAGCAACT	AA	ATATGTTT	TG	ATTTTCAT	TT	GTGTTG	TCAT	660
	GCTTTG	TTAC	TCCTTT	GATG	TACATTA	ATC	AAAAAA	ATTA	TACACTAT	TG	TATATTG	CAA	720
	AGCTAAT	TAA	CTATAACA	AAA	AAGATAG	TTA	ATGCTTT	TGTT	TATTCTAG	TT	AATATAT	AGT	780
	TAATGT	CTTT	TAATATTT	TG	TTTCTTT	AAT	GTAGATT	GGG	CAATTAC	ATT	TTGGAGG	AAT	840
	TAAAA	AAATTA	TGAAAA	AGCA	AATAATT	TCG	CTAGGCG	CAT	TAGCAGT	TGC	ATCTAGC	TTA	900
	TTTACAT	GGG	ATAACAA	AGC	AGATGCG	AATA	GTAACAA	AGG	ATTATAG	TGG	GAAATCA	CAA	960
	GTTAAT	GCTG	GGAGTAA	AAA	TGGGAC	ATTA	ATAGATAG	CA	GATATTT	AAA	TTCAGCT	CTA	1020
	TATTATT	TGG	AAGACTAT	AT	AATTTAT	GCT	ATAGGAT	TAA	CTAATAA	ATA	TGAATAT	GGA	1080
	GATAAT	ATTT	ATAAGA	AGC	TAAAGAT	AGG	TTGTTG	GAAA	AGGTATT	AAG	GGAAGAT	CAA	1140
	TATCTTT	TGG	AGAGAA	AGAA	ATCTCA	ATAT	GAAGATT	AATA	AACAAT	GGTA	TGCAAAT	TAT	1200
	AAAAA	AGAAA	ATCCTCG	TAC	AGATTTA	AAA	ATGGCTA	ATT	TTCATAA	ATA	TAATTTAG	A	1260
	GAACTTT	CGA	TGAAAGA	ATA	CAATGA	ACTA	CAGGATG	CAT	TAAAGAG	AGC	ACTGGAT	GAT	1320
	TTTCAC	AGAG	AAGTTAA	AGA	TATTAAG	GAT	AAGAATT	CAG	ACTTGAAA	AC	TTTTAAT	GCA	1380

**EP 0 786 519 A2**

	5	10	15	20	25	30	35	40	45	50
1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33
34	35	36	37	38	39	40	41	42	43	44
45	46	47	48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63	64	65	66
67	68	69	70	71	72	73	74	75	76	77
78	79	80	81	82	83	84	85	86	87	88
89	90	91	92	93	94	95	96	97	98	99
100	101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120	121
122	123	124	125	126	127	128	129	130	131	132
133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154
155	156	157	158	159	160	161	162	163	164	165
166	167	168	169	170	171	172	173	174	175	176
177	178	179	180	181	182	183	184	185	186	187
188	189	190	191	192	193	194	195	196	197	198
199	200	201	202	203	204	205	206	207	208	209
210	211	212	213	214	215	216	217	218	219	220
221	222	223	224	225	226	227	228	229	230	231
232	233	234	235	236	237	238	239	240	241	242
243	244	245	246	247	248	249	250	251	252	253
254	255	256	257	258	259	260	261	262	263	264
265	266	267	268	269	270	271	272	273	274	275
276	277	278	279	280	281	282	283	284	285	286
287	288	289	290	291	292	293	294	295	296	297
298	299	300	301	302	303	304	305	306	307	308
309	310	311	312	313	314	315	316	317	318	319
320	321	322	323	324	325	326	327	328	329	330
331	332	333	334	335	336	337	338	339	340	341
342	343	344	345	346	347	348	349	350	351	352
353	354	355	356	357	358	359	360	361	362	363
364	365	366	367	368	369	370	371	372	373	374
375	376	377	378	379	380	381	382	383	384	385
386	387	388	389	390	391	392	393	394	395	396
397	398	399	400	401	402	403	404	405	406	407
408	409	410	411	412	413	414	415	416	417	418
419	420	421	422	423	424	425	426	427	428	429
430	431	432	433	434	435	436	437	438	439	440
441	442	443	444	445	446	447	448	449	450	451
452	453	454	455	456	457	458	459	460	461	462
463	464	465	466	467	468	469	470	471	472	473
474	475	476	477	478	479	480	481	482		

55

	TATTTACAC	AGCTTCATTA	ATAAAACGAA	ATTGCTTCAA	CCCGCTTCAA	CTTCAACTGG	3300
	CTTCAACTTC	AGCCTACTTC	ATTCAATAAC	AAAACGAATC	CGCTTCATCC	AAAATCAACC	3360
5	ATTCTAACGC	ACATATTCAA	ATATAGCAGC	TGCACCCATG	CCGACACCAA	TACACATCGT	3420
	AACCATGCCG	TAACGGCTAT	CGGGACGTCT	ACCCATTTCA	TTAAGTAAAC	GCGCGGTTAA	3480
	CATTGCGCCT	GTAGCACCTA	ATGGATGACC	TAAAGCAATA	GCGCCACCAT	TCACATTTCG	3540
10	ACGTGATATA	TCTAGACCTA	CTTCTTTAAT	AGATGCAATC	GTTTGAGAAG	CAAATGCTTC	3600
	GTTCAATTTCG	ATCAAATCAA	TGTCTTCAAC	AGATAGATTG	CTGAGTGACA	ATACTTCAGG	3660
	AATCGCATAT	GCAGGCCCAA	TACCCATAAT	TTTCGGGTCA	ACGCCTACTG	CCTTAAAACC	3720
15	AACGAATCGT	GCAATAGGTG	TCACGCCGAG	TTCTTTCACT	TTATCTCCAG	ACATTAAAAC	3780
	TACAAATCCT	GCACCATCAG	AAAGTGGGGC	AGATGTTCCCT	GCAGTCATAG	TGCCGTCAGC	3840
20	TTTAAATACT	GTACGTAATT	TGGCTAATGC	CTCCATCGTG	GTGTCAGGGC	GTATAAATTC	3900
	ATCTTGGTCA	AAGATATTTG	TGTGTACTTT	TGGTCCTGCG	TTGTATATT	CAACTGAGTT	3960
	TACTTGATTT	GGAATAATTT	CATCTTTGAA	CCGACCATCA	CGTTGTGCGT	CATAGGCACG	4020
25	TTGATGACTT	CTGACAGCAT	AAGCATCTTG	ATCTTCGCGT	GATACGTCAA	ATTGGGATGC	4080
	TACATTTTCA	GCAGTTAAAC	CCATAGGATA	TGACGCACCT	ATATCATCAT	ATTGTAAGGT	4140
	TGGATTGTTT	GTGGGCTCGT	TGCCACCCAT	TGGTACGGCA	CTCATCAATT	CAACGCCACC	4200
30	AGCTACAAGT	ATATCTCCTT	GACCAGCCAT	AATTTGATTG	GCTGCAATCG	CGATGGTTTG	4260
	TAATCCTGAT	GAGCAGTAGC	GATTCACTGT	TTGACCCGGT	ACCGTGTCAG	ATAATCCGCG	4320
	ACGCAATGCA	ATCGTTCGTG	CAATGTTTTG	GCCTTGTAAT	CCTTCTGGAA	AAGCCGTACC	4380
35	AACAATGACA	TCTTCAATCA	TATTCTTATT	GAATTTTCCG	TCAATACGTT	TCAATACGCC	4440
	TTGTAATACT	TTGGCTGCGA	CATCATCAGG	TCTTTCGTGG	AATAATGCCG	CTTGCTTTGC	4500
	TTTCGCTGCG	GCTGAACGCC	CATAAGCTAC	AATGTATGCT	TCTTGATGG	TTATCATCCT	4560
40	CTCTTAATGA	CTATCTTTTA	ATTACGTAAT	GGCTTACCAG	TTTTTAACAT	ATGTGCAATT	4620
	CTTTCATATG	ATTTTTTAGA	TTTTAGTAAG	TCAATAAAGC	CAATTTTCTC	CAACGATTGA	4680
45	ATGTAACGTT	GATTGATAAA	TGTATTTCTT	GGTAAATCAC	CACCCGCTAA	AATTGTGGCG	4740
	ATATTTAAGG	CAATATGATA	ATCATGGTCG	CTAATAAAAT	GACCCCGTCT	TTGCGCATCT	4800
	AATTGTCCTT	GGATCAATGC	TTTGAAGTCT	TCACCTAAAG	CGATATATTG	ATGTCTAGGA	4860
50	TTCGGAATAT	AGTTTGTTC	TGCTTCATAT	TTCGCACGTT	TGAGCGCAAC	TTGACACGT	4920
	TGTGCTGTAT	TGAAAATAAT	CGTATCTGTA	TCACGTAAAT	AACCATAACG	ACGTGCCTCA	4980

55



EP 0 786 519 A2

	TGTTTGTCAT CAAACTTATG CGATGTGCGT AATATGCGAT CAGCCATTTT TGCAAGGCCA	5100
	CCGCCACTCG GTAATAAGCC AACACCTGCT TCAACAAGAC CGATATATGT TTCACTTGCA	5160
5	GCGACAACAA TAGGTGAGTA AAGTACAAGC TCACAGCCAC CGCCTAAGGC ACGACCTTGA	5220
	ACAGCTGTGA CTA CTGTTT CAAACTATAC TTCAAACGAT TAAAGCTATA ATGTAATTTA	5280
	TCAATTGATT GTGCAACGAC ATCATCTACA AGACCGTCTT CATGCGCCTT TTTCATTAAG	5340
10	AAAAGGTTAG CACCCACACT GAAATTGTTA CCATCTGCAT AAATAACCAT ACTTGTGTAA	5400
	TGGTCATTTT CCAGTAAATC AATCGCATCA ACTAACGCAT CGTTGAATTC ATCGGTAATG	5460
	ACATTATTTT TACTTTGTAA TTTCAGTAAC AGTTGATCAT CATGAGTTAC GGAAAGTTTG	5520
15	GCATCACCTT TATCCCAAAG TTCATCTTTT ACGAAGTGAG AAATAGGTGT TGCATATTCA	5580
	ATGGTCTCAT CTTGTTTATA AAAGCCACCA TCTAAATCAC TAATCCATTG TGGTAAGTCT	5640
	CCAAGTTCGT CTTCCATACG TGTTTTAACA CGTTCGTATC CCATTGCATC CCATAATTGG	5700
20	AATGGACCAA GTTTCAGTT GAACCCCCAG ACAAGCGCAC GGTCTATGTC TCGGAAATCA	5760
	TCGGTAGCTT TAGGTACATT GATAGCAGAG TAATAGAAAT TATTACGTAA TGTCTCCCAT	5820
25	AAAAATAGTC CCGCTTCGTC TTGCGCATTG AATATGGTAT CAAGGTTATG CACTAAGTCT	5880
	TTATTAAATT CATTTAAAAT TGGTAATTGT GGTTCGATA CAGGTACATA ATCTTGTTTT	5940
	TCAACATCGT AAACAAGTCG AGCTTTAGTT TCTTTATCCT TTTTGTA AAA TCCTTGTTTC	6000
30	GTTTTACGTC CGAGTGCGCC ATTGTCAAAC AACGTATTTA CAATTTTGAC ATCATGAAAA	6060
	TAAGGTGTTT CTTCAGGTAC TTGTTGCATG CCTTTAATTA CAGACACTGC AATATCTAAA	6120
	CCGACTAGGT CAGATAGCGC ATATGTACCT GTTTTAGGAC GACCAATCGC TTGCCAGTT	6180
35	AAAGCATCCA CATCTACAAT GCTTATCTTG TGTGCTCGG CGCGATACAT AATATCATTC	6240
	ATTGTTGCG TGCCGACTCT ATTGCGACA AAGCCAGGCA CATCATGAC GACAATGACA	6300
	CCTTTACCTA ACACATTTTG CGCGAAATTT TTTACATCTA ATATAATAGA TTCCTTCGTG	6360
40	TGTGACGTAG GTATTAACTC CACTAATTTT ATAATACGTG GTGGGTTAAA GAAATGTAGA	6420
	CCAAAGAATC GTTCTTGATC CTTCTCGTTA AATGCTTGAG CAATCGCATT AATTGGAATA	6480
45	CCTGATGTAT TTGTAGCGAA TAAAGCATCT TCTTTAGCAT GTTGTAGAAC TTGTTGCCAA	6540
	ACAGCATGCT TAATTTCAAT ATCTTCTTTG ACTGCTTCGA TATATAAATC AGCATCATCA	6600
	TTTACCAAGT CATCATCAAA ATTACCATAT GTTAAATGAC TCGCTAGATT TAAGTCGAAT	6660
50	AGTAGCGGCC GTTCTTATC TGTAATTTTA TCGTAAGATT TTTTCGCAAT GAGATTTGGA	6720
	TCGTTTTTGT CCACTACAAT ATCTAATAGT TTTACTTTAA GTCCAGCATT CACAAAAAGT	6780

55

	GTGATTCCTC CAATTTAGTT GAGGATAAGA TAACCATTA AATAATTGGA ATAACGTTGC	6900
	TATTTTATAA AATTAATTAA GTATCTTTGA CAGTCATCTT AGCCTCTTAT TTAAGGAAAA	6960
5	AGCTTTATGC TTAATAAAG TCTTTTTTAG TGAAATTAAT GCATCTCATA TAATTATTTG	7020
	CTATTTATAC GAAAGCAGAA TCTCCAGTCA AAGCGCGTCC AATTACTAAG GCATTAATTT	7080
	CATGTGTACC TTCGTACGTG TAAATCGCTT CTGCATCAGA GAAGAAACGT GCAATATCAT	7140
10	AATCGTCAGC TAGTATGCCA TTACCACCTG TAATACCGCG GCCCATAGCT ACTGTCTCAC	7200
	GCAACGTAA GGCATTCATC ATCTTCGCCG TTGAAGTTGC AACCTCGTCA TATTCACCAT	7260
	GTGCTTGCAT ATTAGCTAAT TGAGCACATG TTGCCATTGC TTGAGCTAAA TTACCTTGCA	7320
15	TCATTGCTAG CTTTCTTGT ATTAAGTATG ATTTACTAAT TGGTTTGCCG AATTGCTTAC	7380
	GCTCAGTGAC ATAATCTAAT GTGGCACGTA AAGCGCCAGC CATACCACCT GTAGCCATAT	7440
	AAGCAACGCC TGCTCTCGTT GAATAAAGAA TTTTGGCAAT ATCTTTAAAG CTTGTTATGT	7500
20	TTTGTAAGCG ATCCGCTTCA TCTACTTTGA CATTAGTTAA TTTAATTAGG GCGTTAGGAA	7560
	CAATGCGAAG TCGATTTTA TTATCAATGA CTTCAATATC GACGCCATCT TGTCTGGTC	7620
25	TGACTACAAA GCAATGGGGT TTGCCAGTTT CTTTATTTAC TCGAATACT GGAATGACAT	7680
	CAGATACATG TGCACCACCA ATCCATTTCT TTTACCAATT GATAACCCAA GTATCGCCTT	7740
	GGCGTTCAGC GACTGTTTCA AGACCTCCCG CAACGTCCGA ACCGTGTTCT GGTTCAGTTA	7800
30	AAGCAAAGCA TGTACGCAGT TCATGTGACT GTAATTTAGG TACATATTTT GCAATTTGTT	7860
	CTTTGCTACC TCCGAAATAG AAAGTGTTAT GCCCTAAACC TTGGTGAACA CCGAGTAGGG	7920
	TAGCTAAGGA AATATCAAAT CGCGCGAGTA GGTAAGACAT GAAAACTGA AATAGTTGAC	7980
35	TAGGCATTTT GCGGTTTGA CGATCCTTGT AAAGTAATGG ATTGTTAAAA TAATTTAATT	8040
	CTCCAGATC TTTAAAATAG TCCTCGGGTA CAGTAGCGTC TATCCAATGT TGATTAATAT	8100
	TTTCACGGTA CTTACTTTCT AGCAATGAAT CTAATTGTTG TAAAAATTCG ACTTCACCGT	8160
40	CTGTTAAACC TTTAGCAATA CTAAGTACAT CTTCAGGAAA TAATGTTTTT AAGACCGTTT	8220
	CTTTTTCAAA TGTCATATAA ATTCCTCCTA AAAATAATAT GAATACTAAT GTGAAATGCA	8280
45	TTTAATTCAA AAACAACACG CTTTATTTGT AAACGCTTAC ACTAAATGTC AAAAATTTTT	8340
	ATCACCTTTA AAGTGTGTTG GAGACTTTGT CATTATCAT TTGTCGAATC GCAAGTTTAT	8400
	CTGGTTTCTG CGTACTGTTT AACGGCATAT GTGTCACTGG TACATACATT CTTGGGACTT	8460
50	TATAACCTGC TAAACGACTT CGCATATGTT GATTTAAAAT TTCAGCGTAA TGAGGTTTAT	8520
	CTTCGCGAAG TATAATGGCT GCAGCAATTG ATTCACCATA TTTTGGATGA TCATAGCCAA	8580

55

	AGACATTTTC GCCACCAGTT ATGATTAAAT CTTTTTTGCG GTCAATAATA AATATATCGC	8700
	CATCGTTGTC CATCTTCGCT AAGTCACCAG TTAATAAATA TCGACCATGA AATGCTTTGG	8760
5	CAGTCTCTGC TGGTTTATTC CAATATCCTG GCGTGACATT TTTAGCCTTA ATTGCAAGTT	8820
	CGCCAATCTC ACCAGTAGGT ACTTCCTCAC CGTTATCATC AAGGATACGT GCATCAACGA	8880
	ACATGACTGC TTTACCAATA CTCATTGGCT TACGTTTTGA ATTTTCCGGT GTATTAAACAA	8940
10	GTACAAGAGG TGCTTCAGTT AAACCATAGC CGTTAATAAT GTTTATGCCA TATTGTTTAA	9000
	AAGCTGCTTG GATACTTGGT AATGGTTGTG AACCACCTTG GATGATATAA TCCATAGCTC	9060
	TAAATTTTTC AGGATTAAAA TTAGTAGCAC GTAGCGTACT ATAATACATT GTCGGAATCA	9120
15	TGATAATAAA TGTAGGGTGA TATTGTGCAA TCATGTCATT CAATTCTTCG CCGTTAAAGT	9180
	AACGTTGAAG AATAAGTGTG CCACCTGACA TTAATACTGG TAATACAGTA TCGTTAAACC	9240
20	CTAAAACATG GAACATTGGT GTTGATACAA TCGTAATATA GTTTGAATTG AACTTATACG	9300
	TCAGCTCTAA GTTTCACCG TTATGAACAA ATGATTCATA TGAGAACATC ACACCTTTAG	9360
	GTGATCCGGT TGTACCACTT GTATAAATTA ATGCTGCAAG ATCTTGTTGGT TCAACAGGTG	9420
25	TTGCTTGAAA AGGTTGGTGA TAATCTGGAT TTACGATTTT ATCATATTGC GCCACATCAA	9480
	TATCCATATG CAATAAGTTT TGGTCAATAT CGGTGAGTGA ACTTAAATGT TTTTCAGCAT	9540
	AGAAGAGCAG TTTTAATTGT GCATCTTCCA CAATGGCTGC AATTTCTTTT GGGTTAAGCC	9600
30	GCCAATTCAA TGGTAAAAAA ACCGCACCTG TTTTAAACA AGCAAACAAT AAATCTAATA	9660
	TTGCAATATC ATTTGGCGCA AAAATACCGA TAACATCGCC TTTTTTAACA CCTTGAGATG	9720
	TTAAATAATG TGCCATATTA TCAGCGCGTG CATTGAGTTG TTGGTATGTC CAAGATGTTT	9780
35	GTTTTGCGTG ATCAATAACG GCAGGCTTGT CATCATCGAA GTCTGAACGC GTTTTTATCC	9840
	AATCGAAATT CATTAGTATA CCCCCTTTAG CTTCACTTTC ATACTTTATG AATTGATTGT	9900
	TTAAGTTGTC CCCATTTTTC TTTGTAAATG CTGGTATCAA TTAATTTTAA ATGATCAGCA	9960
40	ATAATTGGTT TAAAGCCAT TTGATTCAAA ATATCTTTAT GCAAATCAAG ACCTGGTGCA	10020
	ATTTCAATTA GTTTCAGGCC TTGATTGGTG AGTTCGAATA CTGCACGATC AGTAACAAAA	10080
45	TAGATTTCTT GCTCGAGTGA TTGTGAATAT TGTGCATTAA AGTCGATATG GCTCACATCT	10140
	GATACAAATT TCTGGTTTTG TCCTTCAGTT TCAATGTTTA ATCGTTGATT ATGGCATGAG	10200
	ACATGACTGC CAGCTACAAA AGTACCTGAA AAGATAATTT TATTTACAGA TTGCGTAATG	10260
50	TCTATAAAGC CACCACATCC ATTTAGTCGG TCATTGAAGT AAGACACGTT GACATTGCCG	10320
	TATTGATCAA CCTCAGCAAA GCTAAGATAG GCAACTGATA CACCATTGTT ATAAATAAAA	10380

55

	CGACTCCCAA	CGAATCCACC	GAAAATGCCA	ACATCTAAAA	TCGGTTGCAC	ATCATGTTCA	10500
	ACACATTCTT	CATGCAATAA	ATTAGAGAGT	TCATTATTGA	TGCCATAACC	GATGCTAATT	10560
5	GTATCGCCAT	AAGTTAAAAA	CTGAGCAGCA	CGTCGGAGAA	TCAATTTGCG	ACTATTAAAA	10620
	GGTAATGCGG	GTTCAAGTAT	TCCATCAATT	CGTTCTTCTC	CAGACAAGGC	TGGTAAATAA	10680
	TGACTTTGAA	TTACTTGGCG	GTGATTCTTT	TCATCTTCTG	TGACGTATAC	ATAATCGACA	10740
10	AGATTTCTTG	GGATAACAAC	TTCAATTCGGT	TTTAGTTGAT	AGTCGTCAAC	TAAAGCTTTA	10800
	ACTTGATACAA	TAACTTTCCC	ATGATTGGCT	TTGCGGTTTA	ATGCGACATG	ATAACACTCG	10860
15	CTCAAGTACG	CTTCTTGAGT	TAAATAAATG	TTACCTTGTT	GATCTGCGTA	TGTTCTCTCTC	10920
	AGTAGTGCCA	CATCAACGCT	AGGGAATGTG	TAATGTAAGT	ATGTTTCATC	GTTGATGGTT	10980
	ACTAATGAAA	CTAAATCATC	CGTTGTTCGT	GTATTTACTT	TACCGCCACC	GTATCTAGGA	11040
20	TCAACAGCTG	TGTTTAAATCC	GATTTTAGTA	ATAACTCCAG	GTAATAATTG	ATTACTCTGA	11100
	CGATAATGAG	TTGCAATGAT	ACCTTGTTGGT	AAAAAATAAG	CTTCAATGTC	ATTATTTTTC	11160
	ATTGCTTGTG	CCGTTTTGGA	AGAAGCCGTT	AAAATACTCA	TAATGACACG	TTAATCATG	11220
25	CGACGTTCTA	TAAAATCATC	TAAATCCGGT	GCGGCACCTA	AACTATGAAT	ATCATTGCGT	11280
	AATATAAACG	TTAAATCATT	GGGCGTATGA	TATGTGTCAT	GTTGCGCTAA	CACAGCACGT	11340
	AGAACTTCGG	CGGGTAAGTT	GGCTACAGCT	AATGCTGGTA	AACCAATCAC	ATCACCATCT	11400
30	TTAATGATAT	GTTGTAAGTC	GTGCCATGTG	ATTGTTTTCA	AGCAAGTCAC	CTCCATCACA	11460
	TTTGATAAAA	TATAGCGTTT	TTACACTTTG	TGTAAACCCT	TaCAAGAAAT	ATAACATAAC	11520
	GACGTTTAAA	ATCAATTAGA	AATATCTTTT	TATTCTGATA	ATAGACACAG	TATAGACACA	11580
35	TTTTGATGGT	CGATAACAAT	TGTAATATCA	AGGGTTTGTA	ATGAATTGAA	TATCATTAAA	11640
	ATACTTATAT	AAAAATATTG	TTCGGAATAT	AAAAAGTTAA	ATAGGTTTTG	ATTTTTAAAT	11700
40	ATGAAATACA	AAGTGCCCAA	TCGAACAAAG	TATTTATATT	AAAATATGGA	AAATCCATCA	11760
	ATATTAAATT	AAAATAGTTT	TATTATGAAA	AGTGAAAGTA	GGTAAGTCTA	TGGAAGGTCT	11820
	TAATCATCGA	AGAAATACAG	AAAAAGAAGA	GACAACACAA	ACGCAATCaG	TTGCACCTAA	11880
45	TACAGGTGAA	GAGGGGATGT	CATCAGCAAG	TACACAATCA	ACTAAGACGT	CCGACATACA	11940
	TAATGAATCT	ATCGATAAAC	AAATGGAAGC	TAAAGCGCAT	GAAACAGCGC	AAAATACAGA	12000
	TTTAAAAAAC	GAAGCAAGAA	GTTTATTTGA	TAATGCAACC	AAATCAATCG	GTAGACTAGC	12060
50	GGGCAATGAT	GAAAGCTTAA	ATCTTAATTT	AAAAGATATG	CTTTCTGAAG	TATTTAAGCC	12120
	GCATACTAAA	AACGAAGCAG	ATGAAATATT	TATAGCGGGT	ACTGCTAAAA	CTACGCCAGC	12180

55

	TTTCACAGTA ACATTTATTG GATTATGGGT CATGGCAGCA ATTTTAAATA AACTAACGC	12300
	GATTCGGGT CTCATTTT TAGGGGCTTT AACAGTACCA TTATCGGGT TGTCTTCTT	12360
5	TTATGAATCA AATGCGTTT AAAATATTAG CATTTTGTAA GTTATTATCA TGTCTTTAT	12420
	TGGCGGCGTA TTTTCATTAC TAAGTACGAT GGTATTATAT AGATTGTGCG TTTTAGTGA	12480
	TCAATTCGAA AGGTTTGGTT CTTAACATT TTTCGATGCA TTTTAGTAG GATTAGTTGA	12540
10	AGAACTGGA AAAGCACTCA TTATTGTTT TTTGTCAT AAATTGAAA CAAATAAGAT	12600
	TTTGAATGGA TTATTAATCG GTGCTGCTAT TGGTGCAGG TTCGCAGTT TTGAATCAGC	12660
	AGGTTATATT TTGAATTCG CTTAGGAGA AAATGTCCCA TTATTAGATA TTGTCTCAC	12720
15	ACGTGCGTG ACTGCGATTG GTGGTCATTT AGTTTGGTCA KCGATTGTG GTGCTGCAAT	12780
	AGTTATTGCG AAAGAACAGC ATGGCTTTGA ATTCAAAGAT ATTTTGTATA AACGCTTTT	12840
	AATATTCTT TTATCAGCCG TTGTTTACA TGGCATTG GATACATCTT TAACTGTACT	12900
20	TGGCAGTGAT ACGTTGAAA TATTTATTTT AATCGTTATT GTGTGGATAC TTGTATTCaT	12960
	TTTAATGGG GCAGGTTTAA AACAAGTGAA TTTACTGCAG AAAGAATTTA AAGAACAACA	13020
25	GAAAAAGTA GACGAATAAT AATTAAAGCT TATGTTGCTC ATATGTTTGT GACATAAGCT	13080
	ATTTTATAA TTTGTCTTTA AAAGAGTGA ATAGGAATAC TTTTGGAGT TAAAAAGTG	13140
	TTtCACGTTA AACAAATAGT GACAATTAGA TTTATATAA ATGAACATGA TTCACTGAAA	13200
30	GTATGTAATA ATCATTTTAT TGAAATTCAT CAAACAGAAA TTAATACAAT CATATAAGCA	13260
	AATTAAACCA CGCCATAATC ATATTGGATG ACTTCGGCGT GGTTTTTATA GTTGAAGCAG	13320
	GGCTGAGACA TAAATCAATG TCCCACTC CTTATCGTT CAATCGTTGT TCGATAATCG	13380
35	ATTAAATAGA TACCTTCAGG GTTACTTTA TAATTTTAA CCTTAGAGT AGCAGCGACT	13440
	ATTGATCGT TGTAAGCAAT ATAAGTTT GGTACATCTC GACTTGATAA TTTAATAATA	13500
	TCATTAGAAA TATTGTGACG TTCCTTAACA TCTACAGTAT GATTCAATTG ATTAATTAAA	13560
40	TCATCGACGT TGCTATTATT GTAGTCTCCT TTATTAATAG CACCATCTT TTTATATGCT	13620
	TGATTAAAGA AATAACCTGT ATCTCCACGA GGAATTGTT CGAAACTATA CATCGTTGCA	13680
	TCCCATGCAG AACGGTCTT TAAGTAACCT TCTATGTCAT CAACACTTT AATGTCGATT	13740
45	TCAATATTG CTTTTTAGC ATCTGATTGT AACTTTCG CAATTTTCGA TAGCTCTGGA	13800
	CGACCGTCAT ACGTAATTAA CTTAATTTT AAAGGGTGT CTTTGTATA ACCATCTTTA	13860
50	GCTAATAACA TTTTGCTTG TTCGATATTT TGTGTTGTTA ACTTAGGTT TTTAATATAT	13920
	GGAATTTTAT CATTAAATGG ACTCGTTGCA GGTTTCGCAT AACCTTGATA AATATGATCT	13980

55

TTATTAGTAT GATTATACAT AAGTaAGAAG TTCTAAAn

14078

## (2) INFORMATION FOR SEQ ID NO: 192:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 486 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

TGAAAAC	TAA AGT	TTTTCTA	ATGCGT	GA	CT AAAATTAGTA	ATAATTAAGT	TCTCATGATA	60
ATAGGTATTTT	TTGAAAAATG	GAGGAGTCTA	TAAATGGGTA	AAAAAATGGG	TCTAGGTTTA			120
TCTATTGCAT	TGGTTGTTAT	TGGTATTGCC	GTTGTATGTT	TAATGATTTT	TTCTAGTCAA			180
AAAACGACTT	ATTTTGGTTA	TATGAATAGT	AATACAAATG	CAGAAAAAGT	TGTCAGTGAA			240
AAAGATGGAT	TAGTCAAACA	TAATATCAAA	GTAGAACCAT	CTAATGATTT	CAAGCCGAAA			300
AAAGGAGACT	TTGTAAATTT	AGTTTCTAAA	GATGATGGGA	AGACATTTTA	TAAACAAGAG			360
ATTGTTAAAC	ATGATGACGT	CCCACACGGT	TTAATGATGA	AAATTCACGA	CATGCATATG			420
AATTAATAAA	AAAGCATCTA	TAACGTAATT	TTGAAGAAGT	AGAGTTATCT	TCTTATGCGT			480
TTTAGA								486

## (2) INFORMATION FOR SEQ ID NO: 193:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1626 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

GAGGTCTATA	TACAATTATG	GTTGTTCCAG	TTAAACGAAC	TGATGGCTTT	ATTACTAAGT			60
TTAATAGATT	AATTGAAAGA	CGATTATTAC	GTCATTTTCA	G TAAAAAGGT	TATATCACAT			120
GGGAGGAAAA	TTGATTGTCT	GACATTTTAA	AATGTATCGG	TTGTGGTGCG	CCACTTCAAT			180
CTGAAGATAA	AAATAAACCT	GGTTTTGTAC	CAGAGCATAA	TATGTTTCGT	GATGACGTGA			240
TTTGCAGACG	TTGTTTCCGC	TTGAAAAATT	ATAACGAATT	CAAGATGTAG	GATTAGAAAG			300
TGAAGACTTT	TTAAAATTAT	TATCAGGACT	TGCGGATAAA	AAGGGTATTG	TCGTCAATGT			360
CGTGGATGTA	TTTGACTTTG	AAGGATCATT	TATTAATGCA	GTAAACGTA	TTGTCGGAAA			420

TCGAGTTAAA GAATGGTTAA AACGAACAGC AAGAAAATAT GGTTTGGAAG CTGACGATGT 540  
 CGTATTAATT TCAGCTGAAA AAGGCTGGGG CATAGACGAC TTATTATCAT CAATTGCGAA 600  
 5 TATTCGAGAA AATGAAGATG TGTATATTGT AGGGACAACG AATGTTGGGA AATCTACATT 660  
 GATTAATAAA CTGATTGAAG CTAGTGTTGG TGAAAAAGAT GTAGTAACAA CTTCAAGATT 720  
 CCCTGGAACA ACTTTAGATA TGATAGATAT TCCTTTAGAT GAAACATCAT TTATGTATGA 780  
 10 TACACCAGGT ATTATTCAAG ATCACCAAAT GACGCATTTA GTTAGTGAAA AAGAATTGAA 840  
 AATTATTATG CCTAAGAAAG AAATAAAACA ACGCGTATAT CAATTAAATG AGGCGCAGAC 900  
 ATTATTCTTC GGCGGTCTAG CGCGCATAGA TTATGTATCA GGTGGTAAAC GTCCGTTAGT 960  
 15 TTGTTTCTTT TCTAATGACT TGAATATACA TCGTACTAa AC GGAGAAGG CTAATGATTT 1020  
 ATGGCGTAAT CAACTTGCGG ATTTATTAAC GCCACCTGGA AATCCACAAA ATTTTGATCT 1080  
 TAATGAGGTA AAGGCTGTTA GACTTGAAAC AGGCAAAGAG AAACGCGATG TTATGATCTC 1140  
 20 TGGTCTAGGC TTTATAACTA TAGGACCAGG GGCTAAAGTA ATCGTTCGTG TTCCTAAAAA 1200  
 TGTTGAGGTT GTATTAAGAA ATTCTATTTT ATAAGGTGaT TAAAAAAATG AAATTGcAG 1260  
 TTATAGGAAA TCCTATTTCA CATTCCTTGT CGCCCGTTAT GCATAGAGCA AATTTTAATT 1320  
 CTTTAGGATT AGATGATACT TATGAAGCTT TAAATATThC CAATTGAAGA TTTTCATTTA 1380  
 ATTAAGAGAAA TTATTTGAA AAAAGAATTa GAAGGCTTTA ATATCACAAT TCCTCATAAA 1440  
 30 GAACGTATCA TACCGTATTT AGATTATGTT GATGAACAAG CGATTAATGC AGGTGCAGTT 1500  
 AACACTGTTT tGATAAAAGA TGGCAAGTGG ATAGGGTATA ATACAGATGG TATTGGTTAT 1560  
 GTTAAAGGAT TGCACAGCGT TTAnCCAGAT TTAGAAAATG CATACATTTT AATTTTGGGC 1620  
 35 GCAGGT 1626

## (2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:  
 40 (A) LENGTH: 635 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

AGGGTTAATT GTCGGTTTAA TTGCAATGAA TAAGTTCCAT GTATTAGCTG GCTATAGAGC 60  
 50 GAAATTCATC TTAATGGTGA TTTTAACTAT GATGGTCTTC GTACTTATTA ATACGTATTT 120  
 ACTAAGACAG GTAAAATCTA TCGGTATGTT CTTAATGATT GCTGCATTGG GTCTATACTT 180

55

GTCTTATATC GATAACATGT TCTTCAATTA TTAAATGCA GAGCATCCTA TAGGCTTGGT 300  
 GCTAGTAATA TTAACAGTAC TTGTGATTAT TGGCTTTGTA CTGAACATGT TTATAAAACA 360  
 5 CTTTAAGAAA GAGAGATTAA TCTAATGTTG ATGAATAGCG TGATTGCTTT AACTTTTTTTA 420  
 ACAGCATCTA GCAATAATGG CGGACTTAAT ATTGATGTGC AACAGAAGA GGAAAAGCGA 480  
 ATCAATAATG ATTTAAATCA ATATGATACA ACGCTATTTA ATAAAGACAG CAAAGCGGTT 540  
 10 AATGATGCGA TTGCTAAGCA GAAAAAGAA CGACAACAAC AAATAAAAAA TGATATGTTT 600  
 CAAAATCAAG CGAGTCACTC GACTCGCTTG AATGA 635

## (2) INFORMATION FOR SEQ ID NO: 195:

15

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13715 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

25 CTGAAATGGG TATTATTTGT CTTCTTCATC ATAAAGTAAT AAAGATTGTT CATCATTGCG 60  
 ACGTTGCCAA TTTTCATTG GCGATCTTGG AACGTATAAT AATTGCCTAG TAGCATACGC 120  
 TTTGATTGAA ACATACAAGT CATTCCTTGA ACTTGCACCT CAATTTCCAC ATTTGAATTT 180  
 30 TCTTTTATAT TAATAATTC ATCCAAATTC AGCTCACGTG CTAAGACAGC TCTTGATGCG 240  
 CCTCTTTTAC CCCAGTAATT ACATTGAAAA TGATTAGTTA CTAACGTCTC TGCATTCCAA 300  
 TGAAGTGGTA TTGGATTTTC TTGCGCCTTC ACATACATTA CTAAGTCTGG ATCCCCGAAA 360  
 35 ATAATTCTGT CAACTCGTAT TTCATGTAAA AAATTAATAT AATCTTCTAC AGCATCTAAA 420  
 TGATAATTAT GAAATAATCC ATTCACTGCC GCATATACTT TTTTATCGTT TTTGTGAGCT 480  
 AATGCGACAG CCTCTGTCAT TTGTTGTCTA TTGAATTCCC CTGGAAGTCT TAAACCAAAC 540  
 40 TTTTGCTCGC CAATTACAAA AGCATCTGCA CCTAAATCAA TAAGTGTTTC CATATGGCTT 600  
 AATGACTTGG GTGTGACAAG TAATTCTGTC ATAGTCATTC TCCTTTAATT GAAATCGCTA 660  
 45 ATCCATCGTC TATATTTAAA AAATTCGTTG TATATCCTGG TTGCTTTATT AACCCTCAT 720  
 TATAATCTTG AACCTTTTTA ACCATTTGTC TTACATTTCT CGATCTAACA ATCCCAATAT 780  
 CCGATACAAA ACCGTGATAT AAAACATTAT CTGTAATTAC GAGACCTTGG TGCTTTAAAA 840  
 50 GTGGTGATA TATTTCAAAA AATTTCTTTG ATTGCGCTTT TGCTGCATCA ATAAATATCA 900  
 TATCATAAAC TTGTCAATTT ACATTTTCAA ATTGCTCTAA AGCATTACCT TCAATAATTC 960

55



## EP 0 786 519 A2

	CATTACGCTC TATCGTTGTG ACATGAATGT CATCAGATAT AGAAGCGAAT TGCATAGAAC	1080
	TATAGCCGAT TGCTGTACCA ATTTCTAAAA TATTTTAAAC ATTATTCATA CGAATTAATT	1140
5	GCTTAATTAA ATCTAATGTT AAACGATCTA CAATGGGCAC TTCATTTACC TCGGCAAATT	1200
	CACGCAAAAC TTCGATTGAA CTATTTTGAT GTTGATGTAA ATCTATTAAA TATTTTTTAT	1260
	TTAGGTCATC CATGTTTTAA ACTTCCTTTA TGTAATAATA GTCAATATGA TTATGACAAT	1320
10	AAAATAAATC AGCCTTCACA ATTGATTATA ATTTTGCCAA CCAATTAAAT GACTGATTTT	1380
	GTGTTAGACG CAAAGCTATT TTATTTATAG AAGCGAATCA TTCATATAAA ATTTAACTTT	1440
	AGATATTTTA CCATATTTTC AATAAAATTA TAAGCGTTAA TTATTTATAC ATTGCTTGAC	1500
15	TTAAAAATA CTCTTGCCCTC CCCATCTTTA AGGTTAGCAA GAGTAAAATC TTTTAAATTA	1560
	TTCTTCCATT TCAGTATTTA CAACTTCTTC AATCATGTCC CATTCTTCAT CAGTTTCGAT	1620
	TGGTACTAAC TTACCACCGT CACCTGACTC ATCTGGTTCA TTGATCATTG GTACAAGCTC	1680
20	AATCATATCG TCTTCATCTG ATTGAGCACC TTCTTCAGCT AAGATAACAT ACTCTTTTTT	1740
	GAATTCAGGA TGATAAAATT CTAAACTTTT TCGGTATAAA ACTTCATTTT CCTCTTCATC	1800
25	GAATAAAGTT AATAATTCTT CTTCGTTATT AATTTCTAGT TGTGAATCAT GATTATGTTT	1860
	AGTCATAGTA AAATCTCCTT TTAATGTAGT GAATCTAAAT AGCCTTGTA AATAAATACC	1920
	GCTGCCATTT TATCAATCAC TTGTTTTCTT TTTTGTCTTG AAACATCTGC TTCTAATAAT	1980
30	GATCGTTCAG CAGCCATTGT GCTTAATCTT TCATCCCACA TCACAATCTC AATAGAAGGA	2040
	TAAGCTTCTA ATAATTTTTT TTTATATGTT AACGAAGCTT CGCCTCGAAA TCCTATTGAA	2100
	TTATTCATGT TTTTAGGTAG TCCTATTACG ACTGTACCCA CATTATGTTT TTTAATAATG	2160
35	TCTACTAATT GGTCAATACC TAATTCATTA TTTTCTTCAT TGATTCGGAG TGTGTCTAAT	2220
	CCTTGTGCCG TCCAACCCAT TATATCACTA ATTGCAATTC CTACCGTTCT ACTACCGACA	2280
	TCGAGTCCTA AAATTTTATG TTGTAACATA AATTATTTAT TTTGCTCTTT TAAATAGTAA	2340
40	GAAACAAGCT CTTCCATAAT AACATCTCTA TCAATATGAC GAATTTGATT TCTTGCTTCA	2400
	TTTTGGCGTG GAATATACGC AGGGTCACCT GATAATAAAT AACCTACAAT TTGGTTTACG	2460
	GCATTATATC CTCGTTTCATC TAATGTTTGA TAAACATTAT TTAAAACATC TCTTACATCT	2520
45	TGCGTIGGAA GTTCTTCATA GTCGAATTTT ATTGTTTTAT CAAAGTTTTT CATTGCGAC	2580
	ACTCCTTTAA TTACAAATAT AACTCACTAT CATCATACAA TATTATGGCT TTAAATTATA	2640
50	GATTTTTAAT GTAATCTTTA ATAAAGCTTA ATGATTTTGA GATATTTTCA GGTGTGTAC	2700
	CGCCACCTTG AGCCATATCT GGACGACCGC CACCTTTACC ACCAACGATT GGTGCCATTT	2760

55

## EP 0 786 519 A2

	TCGATACTTT ATCATCAACA TTACTTGCAA GAATGATAAT TGTATCTTGT AGTTTAGATT	2880
	TAAAATCGTC CATTGTCGAG CGAATTGCTT TCGCATTTGG TACATCCACT TCAGTAACCA	2940
5	ATACTTTATA GCCATTGATT TCTTCAACTT GATCTTCAAT ATTACCCATT TTAAGTGATG	3000
	TGATTTCTTT GTCACGTTGC TCTAATTGTT TTAATAATGC TTTTCTTCA TCTTGTAATT	3060
	GTGTAACTT ATCGACTACT TGATCATCAG ATTTCACTTT CAGCTGTGAT TTCATCGTAT	3120
10	TAAATTTCTC TTGAATATCT TCTAAATATA AGAAAGCTGC TTTACCTGTT AATGCTTCAA	3180
	TACGACGCAC ACCAGCTCCT GTACCTGACT CACTTACTAT TTTGAATAAG CCAATTTCTAG	3240
	AAGTATTGCG GACATGAATA CCACCACATA ATTCAATTGA AAATGGTGCC ATATTTACTA	3300
15	CACGCACAAC ATCACCATAT TTTTCACCGA ATAATGCCAT TGCGCCCAT TCTTTAGcTG	3360
	AAGCAATATC CATTTCTTGA ATGTAAACGT CAATACCTTT CCAAATTTCT TCATTTACTA	3420
	AGCGTTCAAC TTGATCAATT TCATCATTAG TCATTGGACC AAAATGAGAG AAATCAAAAC	3480
20	GTAAACGATC TGCTTCTACT AGTGAACCAG CTTGGTTAAC ATGATCACCC AGTACTGATT	3540
	TCAACGCTGC ATGTAATAAA TGTGTTGCAC TATGGTTCTT TTGAATGTCA CGTCGATCAT	3600
25	TTTGGTTCAC TTCAGCAGAC ACTGTAGCGC CAACATTTAC TTGGCCAAAT TGTACTACTC	3660
	CTTTATGCAA GTTTTGACCA TTTGGTGCTT TGGTTACTTC ACTAACAGCA ATTTCAAAT	3720
	TGTCATTATA AACAATACCT GTATCCGCAA CTTGTCCACC ACTGATTGCA TAAAATGGTG	3780
30	TTTCCGTAA CATGAAGTAT ACTGTTTCAC CCGCTTCAAC TTGTGAAACT TCTTCACCAT	3840
	TGTATATCAA GTGTGTTAGT GTTGTGTTGAG ctGTGCGAGT ATCATAACCA ACAAAGTAC	3900
	TTGCAGATGT AATATTTTTT AATACTTCAC TTTGAACTTG CATTGATTGA GAATTTTGAC	3960
35	GTGCTTGACG TGCACGATCA CGTTGTTGTT GCATTTCTGA CTCGAATGTT GTCATATCAA	4020
	CTTTCAATCC TGCTTGCACT GCTATTTCTT CAGTTAATTC AATTGGGAAC CCATACGTAT	4080
	CATACAATTT AAATGCATCT TTCCCATTA TTTTCAATTGT TGTCGCTTTA GCTTTTTTAA	4140
40	TTAATTCATT TAAAATCGCT AAACCATCTT CTAATGTTTC ATGGAATCGT TCTTCTTCAG	4200
	ACTTTATAAC ACGCTTAATG AAATCTGCTT TTTCTTAAC ATTTGGATAA TATGGTTCCA	4260
45	TAATGTCTGC AACAATATCA ACAAGTTTGT ACATAAATGG CTCATTGATT CCTAACGTTT	4320
	GACTAAAACG AACGGCACGA CGTAACAATC GACGTAATAC ATACCCTCTA CCTTCATTGG	4380
	CAGGTAATGC ACCATCAGAA ATTGCAAATG CAATCGTACG AATGTGGTCA GCAATTACTT	4440
50	TAAATGCCAC ATCTTGTTCTG TTGTTTACTA AATATTGTTT ACCTGATACT TTTTCGATTT	4500
	CATTCAATTAT AGGCATAAAT AAATCTGTTT CATAGTTAGT ACGTACATTT TGAGAAACTG	4560

55

EP 0 786 519 A2

TATGATCTTT ATTATGATTG AATTCATAA ATACTAAGTT CCATACTTCA AGATAGCGTT 4680  
 CATTCTCTCC ACCTGGATAC ATTTCTTCTG CCGGATCGTC TTGTCCATAT GCTTCTCCGC 4740  
 5 GATCATAGAA AATCTCAGTG TTCGGTCTCG AAGGCCCTTC ACCAATATCC CAGAAGTTAC 4800  
 CTTCAATGCG AATAATACGA CTTTCTTCAA GCCCAATATC TTTATGCCAA ATGTTGTATG 4860  
 CTTCCATATC TTCCGGATGA ATCGTAACGT ACAATTTATC TGGCTCCATA CCCATCCATT 4920  
 10 TATCACTCGT TAAAAATTCC CAAGCAAATT CAATCGCTTC TTGTTTAAAA TAATCACCAA 4980  
 TTGAGAAGTT ACCTAACATT TCAAAGAATG TATGGTGACG CGCTGTGAAA CCAACATTTT 5040  
 CAATATCATT TGTACGAATA GCTTTTGTAG AGTTTACAAT TCTTGGCTTT TTAGGTGTTT 5100  
 15 CACGTCCATC AAAATATTTT TTTAATGTTG CTACACCTGA ATTAATCCAT AATAATGTAT 5160  
 CATCATCAAT TGGCACTAAT GGTGCAGAAG GTTCAACCAT ATGTCCTTTT TCAACAAAGA 5220  
 AATCTAGATA TTTTGTCTA ATTTCACTCG CTTTAACTT TTTATCATT TACACATCCT 5280  
 20 ATTTACTGTT TTAAATTAC CATTCCATAA AAATTGATGA CACAGATAGT CGATTTGCAA 5340  
 AACTAGTATA AATCAATATC ATTTTATT ATTAAAAAT AAAAAACGCC CATCCTCAAA 5400  
 25 AGGGACGAAC GTTATCGCGG TACCACCCTA GTTATAATG CAATTCAACA CATTATCAC 5460  
 TTTAATTCGA CTATACAGTT GTGCATAAAG TAGCGTTCAC TAATGTTTGT TGTACTTTTC 5520  
 ACCAACCACT ACATCTCTGA TAAACAAATC ATTAACACT CATCTTTATA CGAATTTAAT 5580  
 30 TCTATTTTAG TTACATTTAC GCTTGTGTC AACGTTCTAT AAAGTCATAC GCGGTGATTT 5640  
 CTCCCATATT AATCATTGGG TCAATTTTAA ACATTGTAGC TTCCGTTAAT ACATTTGTAT 5700  
 CTGTTTTTGT TGAATCAGAC ATAACCTCTT CACTATCATT CGATGACATT GGCGCTTCTA 5760  
 35 CTTGATCATC TATTGTCGTT TGTGAAGCTC CTGTATCATT AgTTGCTGTG TTTTCCAgCA 5820  
 TTTCTTCATC TTCTGAATTA AAATAATTTT TCAACAATGT ACATAATTGT GTTAAACGCG 5880  
 CTTGACCATT TGTTTTCAAT CCAATATCAA ATGCTTCCGG ATCACCAGT AAAACTAAAC 5940  
 40 TCGTTTTCGC TCTAGTTAAA CCAGTATATA ATATCGGTCT TTGTAACATT CTAAAATACT 6000  
 GTTTAACAAT AGGCATGATA ACAATAGGAA ATTCTGAACC TTGTGATTTA TGGATTGATG 6060  
 TACAATAAGC ATGTGTTAAT TCCATCATAT CTTGTTTCGT AAATGTAATT TCATTACCTT 6120  
 45 CAAAATCCAC AACAAGTACA TCTTTATTAA GGGCATTTTC TTTCGCCCAA AAAATACCAA 6180  
 CAATAACTCC TATGTCACCA TTGAATATGT TATCATTTGG CCTATTACA AGTTGTAATA 6240  
 50 CTTTGTCAAC TTTTCTAAAG ACTACATCAC CAAACTCAAT TTCTCGTGTG TCTTTCTTTT 6300  
 TAGGGTTTAA AATATCTTGT AAAACTTGAT TTAAACGTTT AATACCGGCA TTTCTTTTAT 6360

55

EP 0 786 519 A2

	CTACCTTCTC AACAACTGTT GGTATTGGT TTGCCTGACA GTTAATAAAA CTTCTATCAT	6480
	GAAAACGCTG TGTAATATCA ATTTTCTGAC CCAACTTCAT TCGATGTGCT AATTCTATAA	6540
5	TGCTTGAACC ATCTTGTGA CGATATACTT CAGTCAGATT TACTCGTGGT ATAGCTTTTCG	6600
	ATTCAATTAA ATCTTTAAAT ACTTGACCAG GACCTACAGA AGGCAATTGG TCCTCATCAC	6660
	CTACAAATAT CAATTGTGCA TCTAAAGGAA CTGCACTTAA AAATTGGTGG AACAAACCAAG	6720
10	TATCTACCAT AGACATCTCA TCAATGATTA TGAGTCGTGC GTTTATTTCA TTTTCTAATA	6780
	TATCCTCTGG CTTTGTGTCT TGATTCCAAC CTATTAAACG ATGAATCGTC ATTGCTTCTA	6840
	ATCCAGTTGA CTCTTGTAGT CTCTTAGACG CTCTTCTGT TGGCGCTGCT AATACAACTG	6900
15	GATAATCATC ATTGACATAA TCATCATAAT CTAATGATAA GCCATGAATC TCAGCATATA	6960
	ATTCAACAAT ACCTTTAATT ACTGTCGTTT TTCCTGTTCC CGGTCCACCG GTTAATAGCA	7020
	TCACCTTAGA ATTGATAGCC GTTTGCAAAG CTTCTTTTTG TGAAGCTGCA TAGTTCACCTT	7080
20	GATTCGCATC TTCTATTTCA CCAATATGCA TTTGTAAATC TGACTGTTCA ATTTCTGTAA	7140
	GTTCATTTGT ATGCGTCTTT ATTCTGAATA AGTTTGAAC ACTTTTGATT TCaGAATAAT	7200
25	ACAACTTGG AATTGCAACT TGTTCaTTGT CAATAATTAG TCGTTTTTCC TCATTTAAGT	7260
	ATTGCAACAT TTCGTCTAAT TTTTCAGGTT CGATGACCTC TTCATCTTga TAATTTAATA	7320
	CATCAACCGT TAAATCTATA ACAACATTGA TAGGCAAATA TGTATGTCCC TGTTTAATAC	7380
30	ATTCTTCTTC TAACGTATAG AGCAACGCAG CTTTTAATCG TTCATTATCG TTATAAGCGA	7440
	TACCAATATT TCTAGCAAGT TGATCTGCTT TATTAAAACC AATACCTTTA ATATCATAAA	7500
	TCAATTGATA TGGATTTGCA TCTAAAATAG TCAGTGTATC GCCGAGATAA AACTGATAAA	7560
35	TTGCCATTGA AAGTTTAGGA CCAAACCCTA AATCATGTAA ACGAATCATT ATTTTTTCAG	7620
	ATTCTTGATT TGCTGAAATT TGTTCTGCAA TTTGTTTCTG TTTCTTTTAA GATAATCCCG	7680
	AACTTTTTTC TAGCACTGAA TGGTCATCTA ATATATCATT TATCGCATTG TCACCTAATG	7740
40	TATTAACAAT ATTTTGAGCT GTCTTTTTTAC CTACACCTTT AAACAAATCA CTAGATAAAT	7800
	AACTTATAAT TGCTTCTTTC GTTTGTGGCA TTTCTTTTTC AAAAGTCTCT GCTTTTAATT	7860
	GTITACCATA ACGTGGATGA TCAACAACCT GCCCTTAAA TGTGTAGACA TCGCCTTCAA	7920
45	CAATATTCGG AAGAAACCCT ACAACAGTTG GCATTGTATC AAAGTCTTCA TTTGTTTCAA	7980
	TAGTATCTAC TTTAAGCACT GTATAAAAAT TATCACTGTT TTGAAACAAT ATCGCTTCAA	8040
50	CAGTACCTTT GATCATTGAA TAATCAAATA GTGTAGGGTC TGACATGTTA CTCCTCCTCT	8100
	TTCATTTTAG TGAATGTTTT CAGCGCATGC TGACTTAATA AGTGTTTAGG GTCGATAGTC	8160

55

EP 0 786 519 A2

	AAGCCCAAAT	TGTATCTTGC	ATCAACATGA	TTTTTATCAA	TCGTTAATAC	ATGTTTAAGT	8280
	TGAGTTATGG	CTTCATTAAA	CATTTCTAAT	TGACATAATA	CAAGACCATA	TTGAAATTGA	8340
5	ACTTCTGCAT	CTTTGTCTTT	ATCTAGTTCC	GCAGCAGTCA	TTAAATACGG	CAATGCCAAG	8400
	CTTAAATGAT	TCTAACTGAT	TAAACGCCAT	ACCGATCATA	TAATTACAAT	CAACTTGTTT	8460
	AATCTCTGTT	TGTAATGCTT	GTTGATATAA	TTTAATAGCT	TCTTGATAAC	GTGCTGATT	8520
10	ATAATATACA	TTTGCTAGAT	TAAAAAATAC	GACGCCATT	TTCGGATCTA	TTGTnAAAGC	8580
	TTTTTGAAA	AAACGCTCTG	CCTTTyCAAy	CyCATTCgCA	TCAGCAAGTA	CGATmCCaGC	8640
15	ATTAATATAA	TTTTCAATAA	TTGTAGGATT	TTCTTCGATA	TTCCGAACA	ATGCTTGTA	8700
	CGCTTCTTCT	ATTTTTCCAT	TTTGATGTA	TTGATAAAAT	GTTTGTGAT	CTATCATTTA	8760
	CGAACCTCAT	TTCTCATCAA	TTATAACATC	TTGATAAAAT	GTATGTCTCG	AATCACTTAA	8820
20	CAACGAATAA	AATATAATCT	AATATCATCT	TCATTTCATGA	AAAAGCGGA	ATGGAATAGA	8880
	AATGCTTAAG	AACCATTAA	GGTTTATTAT	GTAATGGTTC	TTCCACATTA	GCCACCACTA	8940
	TTATGTACTT	AAAAATAAGA	ATACATAATT	AGATTTCATGC	ATAGGGAGTG	GGACAGAAAT	9000
25	GATATTTTAA	CAAAATTAAA	TTCGTTATCC	CCAACGGCA	TTGCCTGTAG	AATTTCTTTA	9060
	CGAAATTCTC	TATGTTGTGG	TCCCGCCAAT	ATAACATTGT	AGAGCCTAGG	ACATTGTGAT	9120
	GTCCCAGACT	CTATCCTCAT	GAATTATTCT	CATCAAAAAC	TGTCTTTCGT	CATTTTCAAC	9180
30	GTTGAAACTT	CAAATAAGTA	ATTTATTGTT	GCCATTGTTT	ATACAACATA	ATTTAATTGA	9240
	CCTTCATTTT	TGAACACATC	GTCAATTGTT	GCACCACCAA	GACACACATC	ACCTTGATAA	9300
	AAAACAACTG	CTGTCCAGG	TGTGATTGCT	CTTACTGGCT	CAGCAAAAGT	AACACGTAGg	9360
35	CAtGGtCGTT	TTACGTTTTC	ACAAAACTT	TCGTATCTTT	TTGGCGATAT	CTAAATTTAG	9420
	CTGtACATTC	AAAACCTTGA	TCTAAGTCAT	TATCTTCTGG	ATTTACAAAT	GAATAGTCTG	9480
40	AAGCAATTAA	GTAATCACTG	TATAATGCAT	CGTGATGGAA	TCCTTGTTCT	ACATATAAAA	9540
	CATTATCTTT	TAGGTTTTTA	CCGACAACAA	ACCAAGGATC	GCCATCTCCA	CCTATACCTA	9600
	ATCCATGTCT	TTGTCTTATT	GTGTAATACA	TCAAACCACT	ATGTTTACCC	ATTTTCTTAC	9660
45	CATCAAGTGT	TATCATATCA	CCCGTTGTG	CAGGTAAATA	TTGTGATAAA	AATGTTTTAA	9720
	AGTTTTTTTC	GCCGATAAAA	CAAATGCCTG	TAGAATCTTT	TTTCTTAGCA	GTAACAAGTC	9780
	CTTGTCTTTC	AGCAATTCTGA	CGCACTTCAC	TCTTTTCGAT	GTCGCCAATT	GGGaACATCA	9840
50	CTTTTGAAAG	TTGTTGTTGA	GATAATTGAT	TCAAGAAGTA	TGTTTGATCT	TTATTATTAT	9900
	CTACACCACG	TAACATTTCA	ACATGACCAT	CTTCATGACG	ATGTATGCGT	GCGTAATGTC	9960
55							

	TTTCTTTATT ACACATAACG TCTGGATTTG GAGTACGACC TTTTGTGTAT TCATCTAAGA	10080
	AATACGTAAA GACTTTATCC CAATATTCTT TTTCAAAATT AACAGCGTAA TACGGAATGC	10140
5	CAATTTGATT ACACACTTCA ATAACATCGT TGTAATCTTC AGTTGCAGTA CATACGCCAT	10200
	TTTCGTCACT GTCATCCCAG TTTTTCATAA ATATGCCAAT GACATCATAA CCTTGTCTCT	10260
	TTAAGACGTG GGCTGTTACA GAACTATCTA CACCGCCTGA CATACCAACG ACAACACGTA	10320
10	TATCTTTATT TGACAATTAT GACTCCTCCT TAAATTTAAA ATATATTTTA TGAATTTTCT	10380
	CTACAATTGC ATTAATTTCA TTTTCAGTAG TCAATTCGTT AAAACTAAAT CGAATCGAAT	10440
	GATTTGATCG CTCCTCATCT TCGAACATTG CATCTAAAAC ATGCGACGGT TGTGTAGAGC	10500
15	CTGCTGTACA TGCAGATCCA GACGACACAT AGATTTGTGC CATATCCAAC AATGTTAACA	10560
	TCGTTTCAAC TTCAACAAAC GGAAATATA GATTTACAAT ATGGCCTGTA GCATCCGTCA	10620
	TTGAACCATT TAATTCAAAT GGAATCGCTC TTTCTGTAA TTTAACTAAA AATTGTTCTT	10680
20	TTAAATTCAT TAAATGAATA TTGTTATCGT CTCGATTCTT TTCTGCTAAT TGTAATGCTT	10740
	TAGCCATCCC AACAATTTGC GCAAGATTTT CAGTGCCTGC ACGGCGTTTC AATTCTTGTT	10800
25	CACCGCCAAG TTGAGGATAA TCTAGTGTA CATGGTCTTT AACTAGTAAT GCACCGACAC	10860
	CTTTTGGTCC GCCAACTTA TGAGCAGTAA TACTCATTGC GTCGATCTCA AATTCGTCAA	10920
	ACTTAACATC AAGATGTCCA ATTGCTTGAA CCGCATCAAC ATGGAAATAT GCATTTGTCT	10980
30	CAGCAATAAT ATCTTGAATA TCATAAATTT GTtGCACTGT GCCAaCTTCA TTATTTACAA	11040
	ACATraTAGa TACTAAAATC GTCTTATCTG tAATTGTTTC TTCAAGTTGA TCTAAATCAA	11100
	TAGCACCTGT ATCATCAACA TCTAGATATG TTACATCAAA ACCTTCTCGC TCTAATTGTT	11160
35	CAAAAACATG TAACACAGAA TGATGTTCAA TCTTCGATGT GATAATGTGA TTACCCAATT	11220
	GTTCaTTTGC TTTTACTATG CCTTTAATTG CCGTATTATT CGATTCTGTT GCGCCACTCG	11280
	TAAATATAAT TTCATGTGTA TCTGCACCAA GTAATTGTGC AATTTGACGT CTTGACTCAT	11340
40	CTAAATATTT ACGCGCATCT CTTCCCTTAG CATGTATTGA TGATGGATTA CCATAATGCG	11400
	AATTGTAAAT CGTCATCATC GCATCTACTA CTTCAGGTTT TACTGGTGTG GTCCGAGCAT	11460
	AATCTGCATA AATTTCCATG TTTGGACACT CCTCACAATT TTATCAATGT TCCAATAATA	11520
45	GCACCTTACA TACTATTTTT CTACTTTTCT GTTTAACTTT ATTTATAATG TTTTAAATTA	11580
	TATTTTACCA TTTTCTACAC ATGCTTTTCG ATAGGCTTTT TTAAGTTTAT CGCTTTATTC	11640
50	TTGTCTTTTT TATAAATTTT AGTATTTGCA GATATTTTTT TATTTGTAAA ATGTAACGTA	11700
	CTATTATTTT GGTATGAGC AATTTAATAT TTATCTGGTT ATTCGATTGG TATACTTCTT	11760

55

EP 0 786 519 A2

	GTCCCTATTC GAGAAGGTGA AGATGAACAA ACAGCAATTA ATAATATGGT TAATCTCGCA	11880
	CAACATTTAG ACGAATTATC ATATGAAAGA TATTGGATTG CTGAACACCA TAACGCTCCC	11940
5	AACCTAGTAA GTTCAGCAAC TGCTTTATTA ATTCAACATA CGTTAGAACA TACGAAACAC	12000
	ATACGTGTAG GTTCTGGAGG CATCATGTTA CCTAATCATG CTCCATTAAT CGTTGCGGAA	12060
	CAATTTGGCA CGATGGCAAC ATTATTTCCA AATCGTGTCTG ATTTAGGATT AGGACGTGCA	12120
10	CCTGGAACAG ATATGATGAC CGcAAGTGCA TTAAGACGAG ATCAACATGA TGGTGTTTAT	12180
	AAATTTCCAG AAGAGGTTTC ATTATTACAA CAATATTTCTG GCCCTGCTCA CCAACAAGCA	12240
	TATGTTCTGT CTTATCCAGC AGTAGGTAAA AATGTGCCTT TATACATTCT TGGTTCTTCA	12300
15	ACAGATTCTG CACATTTAGC TGCTCGCAAA GGGCTTCCAT ATGTGTTCTGC TGGACATTTT	12360
	GCACCTCAAC AAATGAAAGA AGCTATCGAA ATTTACAAAA CGTTATTTGA ACCTTCTGAT	12420
	GTATTAGACG AACCTTATGT TATTGTATGT TTAAATACAA TCGTTGCTGA AAATGATGAC	12480
20	GAAGCACAAT ATTTAGCTTC ATCTATGGCA CAAGTAATGG TTAGTATCAC TCGTGGCAGA	12540
	ATGCAGCCCG TTCAACCGCC AACACATGAA CTACAAAATA TATTAACGCC GAGAGAATAC	12600
25	GCGATGGCTA TGGAAAGACA GAAAATATCA TTAATAGGTT CAGAAAATAC TGTTCACAA	12660
	AAAATTCAAG ATTTTATGGA AACTTATGGT GAAGTCAACG AAATTATGGC AATAAGTTAT	12720
	ATTTATGATA AAGATATGCA ATTAGACTCT TATCGTCGGT TCAAGAATGT TATAAATCAG	12780
30	ATAAATGAAA AAAACACTTT ATAATGTGAT AAATAAACTA AGTGAAAGTA TGTATCCATA	12840
	ATATTAATAA AAATATACAG TAACAGCATT TTGAATGAAA GATGTCTTTA TTGTTCAATC	12900
	ATTTATTTTA GTAATGATTC AAATTCACTT AAAATyCTAA tGCAAATATG AAAGCGCCCC	12960
35	TTCAcTTTAC ACTGTGTAAG TGTTTATTTG ATGGGGCGCT TTCAAAATAT TGAAAAGCAT	13020
	ATCCzAAAATT TAAAGAAATT TATTTCTCTT TATCTTCATT TTCTTTTTC TCTTCGTTAT	13080
	TCGATCCTGT ATATTCATTT ATCTTATCTT TTACATTTT AACTTGTTC TATCGCTAT	13140
40	TTTTAAATTT TTCTACGCGT CTTTAGCTTT ATCCATAAAA CTCATATTAA TCGCTCCTCT	13200
	TATATTTGAT TAGTTTAATT GAACTTATTT TTTAAGTTA TCAATTGCAT CAGTTATTTT	13260
45	GTTTTTAGCA TTTTCAACAA CTTCCTTTGC TTTaCCAGTC GCTTTATCTT GCTGACCTTC	13320
	TTTTTCTAAT TCTTGTTAT CAGTAACGTT ACCTACTGTT TCTTTAACAT TTCCTTTAAA	13380
	TTGATCGAAC TtACTTTCGT CTGCCATAGT GAAACCTCCT TGGATGTATA TATTTATATA	13440
50	CCACTAAGGA GGTCGCTmm mCAyymyAAT ATGAAGTTT TATGTTATAG TATAGTATTT	13500
	ATACGATTAA ATATAAAACA TGTATCCGTC TAAATCTTCA CTTGTATCTA CATATCCGC	13560

55

TAGTTGTTTT TGCGCAGGTG GTTCTGATTC AATACTTTCA ACAAATGTAA TTGGACCTTC 13680  
 TAACAGTCTT ATAATATCCC CTGCTGAGAT TTCTT 13715

5 (2) INFORMATION FOR SEQ ID NO: 196:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 873 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

15 AAATCCATAA TGTCATGATA ATCTGCATAT GCTTCATATA ATTCAATCAT TGTGAATTCA 60  
 GGGTTATGTC TAGTTGATAC ACCTTCATTA CGGAATACTC TACCAATTC ATATACTTTT 120  
 20 TCAAGTCCAC CGACAATTAm ACGTTTTTAAA TGCAACyCAA TAGCAATACG CATGTATAaC 180  
 GTTGCATCTA ATGCATTATG ATGTGTTACA AATGGTCTAG CAGCTGCTCC ACCAGCAATT 240  
 TGGTGCATCA TAGGTGTTTC TACTTCCAAG AAACCTTTAT TATTTAAATA ATTACGCATT 300  
 25 TCTTGAATGA TTTTACTACG ATTAATAAAT GTACGAGTGC TATCTTCGTT CGTAATTAAA 360  
 TCTAAATATC TTTGACGATA tCTCTGTTCA ATATCCTGTA AACCGTGGAA TTTATCCGGT 420  
 AATGGTCGCA ATGATTTAGT TAGTAGCGTG AATTCTTCG CTTTAACCGA TAATTCGCCA 480  
 30 GTATTTGTTT TGAACATTAC ACCTTCAACA CCAACGATAT CGCCTAAATC AGCATTTTTC 540  
 CATAAATCAA ATTCGTCATC GCCAACTTGA TCTTTACGAA CGTAAATTG AATTTGTCCA 600  
 GCTAAGTCCT GAACGTGTGC AAATCCTGCT TTACCTTTAC CACGCTTAGT CATTAATCGT 660  
 35 CCAGCTATAG CGACATGACT ATCCGCTTCT TTTTCTACCA ATTCTTCTTT AGAATACTGG 720  
 TCCCACTCTT CTTTCAAATC ACTAGATAAA CCTGAACGGT CAAATTTAGA ACCAAACGGG 780  
 TCTATACCAA GATCATATAA TTCTTGTAAT TtTTGACGTC GAACCAACAT TTGGTCATTC 840  
 40 ATTTCTTCTG ACATAACTtT CTCTCCTTTA ACT 873

(2) INFORMATION FOR SEQ ID NO: 197:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 452 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

55



ACCATAATAT GAATGGCTTC AGGATCAAAA TAAAGACCAA CTTCAGTACC TACTTCAGCT 120  
 TTTTGTAGTCG TTTGTATTAC CCATTCTATA CCTTTATTGT CTATACAACA TATTTCATAG 180  
 5 TGGACCCCTC TAAATAACAT AGAATCAACA GTTGCTTTAA ATAATCCTTC TTCAGCTTTG 240  
 ATTAATGATA TATCTTCTGG TCGAATAACG ACTTCTACTT TTTTATTTTC AGGAATACCC 300  
 ATATCGACAC ATTGAAATC TTGCCATATA ATATTACGA CATAATCTCT AACCATGCGC 360  
 10 CCTTCAACAA TATTAGATT TCCAATAAAA TCAGCTACAA ATCGATTAC TGGTTCGTCA 420  
 TaTATATCTG TTGGTGTGCC AAATTGTTGA AT 452

## (2) INFORMATION FOR SEQ ID NO: 198:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2308 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

25 TAGGTTGGGT TCTAACATAC GATAAGCTC AACAAATCAA CACAGCTTTC TTTGTAAAAT 60  
 TGTTTAATAC TGCATTAGCA GAACGTGATT ATTATTTTAA TATAGATGGA ACAAATGCTT 120  
 TTAGATTATT TAATGCTGAA GGTGATGGTG TTGGGGGATT AACAATCGAC AATTACGATG 180  
 30 GTCATTTGTT GATTCAATGG TACTCAAAAG GATTTTATAA ATTTAAATAT GCCATTCTTG 240  
 AAGCGGTTAG AAAAGTATTT GATTATAAAT CTATTTACGA AAAAGTAAGA TTTAAAGACA 300  
 GCGAATATAG TGGTGGTTTT GTTGAAGGAG ATGCACCTGa GTTCCAATT GTTATCGAAG 360  
 35 AAAACTTCAC ATTTTATAAT GTAGACCTTG AAGATGGTTT GATGACAGGT ATCTTTTTTAG 420  
 ATCAAAAAGA AGTGCGCAAG AaATTAAGGG ATCAATATGC CAAAGAACGC CATGTTTTAA 480  
 ACTTATTTAG TTATACAGGT GCTTTTTCTG CAATAGCAGC AAGTGAGGCA TCTTCAACAA 540  
 40 CAAGTGTAGA TTTGGCTAAT CGTTCTCGTA GTTTAACTGA AGAAAATTTT GGATTAAATG 600  
 CTATTGATCC TAAATCCCAA TATATTATG TCATGGACAC TTTTGATTTC TATAAATATG 660  
 CTGCACGACA TGGACATAGT TATGACACGA TCGTGATTGA TCCACCTAGC TTTGCGCGTA 720  
 45 ACAAAAAACG TACATTTTCA GTGCAAAAAG ATTATGACAA ATTAATTAAT GGCGCCTTAA 780  
 ATATCTTATC ATCTGAAGGA ACATTATTGT TATGTACAAA CGCAAGTGTA TATCCATTAA 840  
 50 AGCAATTTAA AAATACTATT AAAAAGACGC TTGAAGAGAG TGGCGTTGAT TATGAATTAA 900  
 CTGAAGTTAT GGGATTACCA AAAGATTTTA AAACGCATCC ACATTATAAG CCATCTAAAT 960

55

TATTGAGAAA AAGAAGGGTG ATAATATTAT GGGATTCAAA AACAAATTAA CATCAAATTT 1080  
 AACAAATAAA ATCGGTAATT CAGTCTTTAA AATAGAAAAT GTTGACGGAA AAGGTGCAAT 1140  
 5 GCCAACGACG ATTCAAGAAT TGAGAGAAAG ACGACAACGT GCTGAAGCAA TTGTAAAGAG 1200  
 AAAGTCTTTA ATGTCATCAA CAATGAGCGT TGTTCOAATT CCGGGTTTAG ATTTTGGTGT 1260  
 TGATTTAAAA TTAATGAAAG ATATTATCGA AGATGTTAAT AAAATTTATG GTTTAGATCA 1320  
 10 TAAGCAAGTT AATAGCCTTG GGGATGATGT GAAAGAAAGA ATTATGTCTG CAGCAGCAAT 1380  
 TCAAGGTAGT CAATTTATTG GTAAAAGAAT TTCAAATGCA TTTTAAAAA TTGTAATTAG 1440  
 AGATGTAGCT AAACGTACTG CTGCAAAACa AACAAATGG TTTCTGTG TAGGACAAGC 1500  
 15 TGTGTCTGCA TCTATTAGTT ACTATTTTAT GAATAAAATT GGAAAAGATC ACATTCAAAA 1560  
 ATGCGAAAAT GTTATTAAAA ATGTCATGTA GGTGCTATAA TAGTTTTGCA ATTTGCAAAT 1620  
 20 TTTACTGAAA CCGGTTTTAA ACGAATTGAA TTAAAGcAT GGTTTTGGTA AAGTTAATGT 1680  
 ATAAACTAA GTTAGyATTG TAATAATATk GAAGATTCTA ACTATACGAA GGAGAAATGT 1740  
 AATTATGGAA CAAAATTCAT ATGTAATCAT CGACGAGAmT GGTATTCACG CTAGACCAGC 1800  
 25 AACAAATGTTA GTACAAACAG CTTCAAAATT CGATTCTGAT ATTCAATTAG AATATAACGG 1860  
 TAAGAAAGTA AACTTAAAAT CAATCATGGG TGTTATGAGC CTTGGTGTG GTAAAGATGC 1920  
 TGAAATTACA ATTTATGCTG ACGGTAGTGA TGAATCTGAC GCCATTCAAG CAATCAGTGA 1980  
 30 CGTCTTATCA AAAGAAGGAT TGAATAAATA ATCATGTCTA AATTAATTAA AGGTATTGCC 2040  
 GCATCTGATG GTGTCGCAAT TGCTAAAGCT TATTTATTAG TTGAGCCAGA CTTAACATTC 2100  
 GACAAAAATG AAAAAGTCAC TGATGTTGAA GGAGAAGTTG CAAAGTTCAA TAGCGTATC 2160  
 35 GAAGCTTCTA AAGTTGAGTT AACTAAAATT AGAAATAATG CAGAGGTTCA ACTAGGTGCT 2220  
 GATAAAGCTG CTATCTTTGA TGCacaTTGG GGGGTGGTAG ATGACCCTGA ATTAATTCAA 2280  
 CCAATCCAAG ATAAGATTAA AAATGAAA 2308

(2) INFORMATION FOR SEQ ID NO: 199:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5559 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

AAGTAATAAA TCGTCTCATT TGGCAACTGA CGCATAATTT CTTTAGCTAC TGTCAAACCT 60

	TTTAATTTTT AGTTTATCAT AACTAAGCAT TGGATTTTAG TATTATGCAC TGTGTTTACC	180
	ATTTTTGTCA TTATAATATT TATTTTAAAT CAGCCCACTA TCATATTGTC ATGTAATCTG	240
5	CTTATTAAAA AAATCCCTTC CAAGTTATTG TGTATCTCCA TTCAATTTAA TTTTGAAAGG	300
	AACATAACWT TTAACTCAA AAGGGATTAA TTTnTAnTCT ACTTCATGGT CTGAACCAAA	360
	GAATGATTTA AACATGTGGA ATGTTGTTTC TCTGTTCAAT GCTGCAATGG ATGTTGTAA	420
10	TGGAATACCT TTAGGGCAAG CATTAAACACA GTTTTGTGAA TTACCACACT GCTGTAAGCC	480
	ACCAGTACCC ATTAATGCAT TTAAACGTTT ATCTTTAGTC ATAGATCCTG TTGGGTGCAA	540
	ATTAAACAAA CGAACTTGCAG AGATTGCTTG TGCACCAACG AaTTTATTAT TTCAGTAAC	600
15	ATTAGGACAA ACCTCTAAAC ATACACCACA TGTACATACAT TTAGATAATT CATAAGCTGT	660
	TTGACGTTTT TTCTCTGGCA TACGTGGTCC CGGACCTAAA TCATACGTTT CATCAATTGG	720
	GATCCATGCT TTCATACGTT TTAAGTTATC GAACATTCTA GAACGATCAA CTTGTAAGTC	780
20	ACGGATAACT GGGAAAGTAT TCATTGGCTC TAAACGAATA GGTTGTTCTA ATTGATCAAC	840
	AATCGCAGAA CAAGATTGTC TTGCACGACC ATTGATAACC ATAGAACATG CTCCACATAC	900
25	TTCTTCTAAG CAGTTCATAT CCCAGACAAC AGGTGTTGTT TTTTCACCTT TAATATTAAC	960
	TGGGTACGT CTAATTTCCA TTAAACAAGC AATGACGTTT AAATTTTCAC GATATGGAAT	1020
	TTCAAATGTT TCTTCATAAG GCTTAGAATC ACTTGATCTT TGTCGTTTAA TAATTAATTT	1080
30	TACTGTTTTT TGTTTCGGTT TAGATTGTGT TTCATGTTGT GGAGTGTTTT TCACTGATTG	1140
	TTCAGTCATT ATTTTTTACC CCCTTTAGAC TTAAGTTGTT AATCACGTTT ACGAGGTGGT	1200
	ATTAACTCA CATCGACGTC ATCATAAGTA AACTGCGGTT TTTCAAATGC GCCTTGGAAT	1260
35	GAGGCCATTG TCGTTTTTAA CCACTCTTCA TCATTACGCT CTGGGAATTC TGGTTTATAA	1320
	TGGGCACCGC GTGATTCGTT ACGGTTATAT GCACCAATCG TAATAACACG TGCAAGTACT	1380
	AACATGTTCC ATAGTTGACG GGTAAAGAAT ACCGCTTGGT TACTCCAAGT TTGAGTATCT	1440
40	TCCATATCAA TATCTTCATA ACGTTTCATC AATTCAACAA TCTTTTATC TGTTTCTAAC	1500
	AGTTTTTCAT TTCACGAAC AACAGTTACA TTTGCTGTCA TAATTCACC AAGTTCACGG	1560
45	TGTAATTTAT ATGCATTTTC TGTACCGCGC ATAGCTAATA ATTTATCAAA ACGTTCTGTC	1620
	TCTTCAGCTT TACGCTTTTC AAAAATACTT TCGTCCATAT CAGTATATGA TCGATCAATA	1680
	TTTGAAATAT AATCAATCGC GTTTGGACCT GCTACTGTAC CACCATAAAT CGCTGATAAC	1740
50	AATGAATTGG CACCTAAGCG GTTACCACCA TGTTGAGAGA AGTCACATTC TCCAGCTGCA	1800
	AATAACCCTT TAATATTTGT CATTTGATCA TAATCTACAT ATAGACCACC CATTGAATAG	1860

55

## EP 0 786 519 A2

	TAAATCTCAA TGATACCACC TAGTTTTACA TCTAACTCAT GTGGATCTTT ATGTGACAAA	1980
	TCAAGATATA CCATGTTTTTC GCCATTTATA CCTAATTTTT GGTAAATACA TACATCGAAA	2040
5	ATTTACGCG TTGCGATATC ACGAGGTACT AAGTTACCAT AATCAGGATA TTTCTCTTCT	2100
	AAGAAGTACC AAGGCTTACC ATCTTTATAT GTCCAAATTC GTCCACCTTC ACCACGTGCT	2160
	GATTCACTCA TTAGTCGCAG TTTATCATCA CCAGGGATTG CAGTAGGATG AATTTGAATG	2220
10	AACTCACCAT TAGCATAAAT AGCGCCTTGT TGGTAAACAA TGGAAGCCGC TGATCCTGTA	2280
	TTAATCATTG AGTTTGTGT TTTACCGAAA ATAATACCAG GGCCACCCGT TGCCATAATA	2340
	ACTGCATCTG AACCAAATGT TTCAATCTCA GCAGTTGTCA TATTTTGTGC AnCGATACCT	2400
15	CTTGCACTAT CATCGTCACC TTTAACTATG CCAAGGAATT CCCATCCTTC ATACTTCGTA	2460
	ACTAATCCAT CTACTTCATA TGCACGAACT TGTTCATCCA ATGCATATAA TAATTGTTGT	2520
	CCAGTTGTTG CCCCTGCATA TGCTGTTCTG TGATGTAATG TACCACCGAA ACGTCTAAAA	2580
20	TCTAATAGAC CTTCATTTGT TCTATTGAAC ATTACGCCCA TACGGTCTAA TAAATGAATA	2640
	ATTTTAGGTG CTGCCTCTGT CATCGCTTTA ACAGGTGGTT GGTTTGCAAG GAAATCGCCA	2700
25	CCATACACTG TATCATCAAA GTGAATCCAA GGAGAATCGC CTTCCCCTTT AGTATTGACC	2760
	GCACCATTAA TGCCACCTTG GGCACAAACA GAGTGCGAAC GCTTTACTGG TACAAC TGAG	2820
	AACAAATCTA CATGTGCACC TTTTCTGCC GCTTTAATTG TTGACATTAA GCCCCTAGG	2880
30	CCACCTCCGA CAACAATAAG ATGTTTCTCT GCCATAAAAA TTCACTCCC CTAAATTTTC	2940
	AATCTATATT TGTTAAATGC GATGTATTAC ATAAAGGCAA TAATTGCAGT AACACCAATA	3000
	TACGAAATAA CTAAAAATAC GATTAATGAA ACCCATGTAA ATACTCGTTG TGATTTTGGA	3060
35	GATTGAAGTC CACCCCAAGT AACTAAGAAT GACCATAAGC CATTTGCAA GTGGAACACA	3120
	ACAGCAATAA TACAAATAAT ATAAATATT GCCCATCCAG GATGTTGCAA TGTTTCGTGC	3180
	ATTAAATCGT AATTCACCTC TTTGCCGTAA AATGCTTTTT GTAAACGTGT TTGCCATAAA	3240
40	TGGATACCAA TAAAGATAAA TGTTAAGATA CCACTCACTC TTTGGAAGAA GAACATCCAG	3300
	TTTCTAAAAA TCGAGTAATG TCCAACATTT TCTTTTGCTG TAAATGCAAT GTGTATACCA	3360
	AACAAACCGT GATATAACAA CGGAATGTAT ATAAATAAAA ATTCTACAAT AATTAGAAAT	3420
45	GGTAATGATT CCATAAAGTT AGATGCCTTA TTAACGCTT CAGCACCTTG TGTTGCTTGG	3480
	TGATTCACTA ATAAATGAAC GACCAAAAAT GCACCTATTG GGATAATACC TAATAACGAG	3540
50	TGAATACGTC TTAGATAAAA TTCATTTTTT GATTGAGCCA AAAGGAGTCC CCCCTGTGAA	3600
	CGAATATTTA ATTTATTGAG CTATTTATAT TAAACGTACG CTTAACCCCC TAAAGTGATA	3660

55

## EP 0 786 519 A2

CGATCACCAA ACTGCATGTC GAACAATGTA ACATTTGGAT TCGATATTTA AAATTGCTTG 3780  
 TGATGATAAA CTTTCTCATT TAGAAAACGC TTCCACGTAC ATTCAAAAAA ATAACTTTGT 3840  
 5 TAACCATATT GTAACATTAT TTCATATATT TTGGGGCATG AGAATGATTC TCACGCCAG 3900  
 TAATTTATTT ATGCAATTGT TCATGTAGGT TCTTTGCGAC GTTTTTCAGGA ATACCTATAT 3960  
 TTTTAAATC TTCAAGTGTA GCTTCCTTCA TTTTCTTGAT TGAACCGAAT GAACGCAATA 4020  
 10 ATAATGTTTT ACGTTTGTTA CCGATACCAT CTATATCATC AAGTATTGAT TTCAAGCCTG 4080  
 TCTTTTGACG TGTTTGCTA TGAAATGTGA TTGCGAATCT GTGAACCTCA TCTTGATAC 4140  
 GGTGCAACAA ATAAATGCC TGGCTATTTT TCTTCAGTGG TACAATTTCT GCACTAGCGC 4200  
 15 CATATAATAA TTCAGATGTT TGGTGTATAT CATTTTTCTG CAAACCTGCA ACAGGGATAT 4260  
 CAAGACCTAA TTCGTTTTGT AGCACATCAA TAACCCCGTT CATATGTCCT TTACCACCAT 4320  
 CTACTATTAT TAAATCAGGT AATGGTAATC CTTGTTTTAA AACGCGAGAA TATCGTCGTC 4380  
 20 TTACTACTTC TCTCATTGAT TTGTAATCAT CTGGACCTTT AACCGTTTTG ATTTTATACT 4440  
 TTCTATAATT TTTCTTATCT GGTTTACCGT CGACAAATGT AACCATTGCT GACACTGGAT 4500  
 25 CCACACCTTG AATATTAGAA TTATCGAATG CTTCAATTCT AATTGGTGTT TGAATTCCCA 4560  
 TTTGTGTCC AAGTTCTTCA ATAGCTTTAA TCGTTCTGGA CTCATCACGT GATATTAATT 4620  
 CAAATTTATT ATTTAAGGAT ACTTTAGCGT TATGTGCAGC TAGGTCAACC ATATCTTTTT 4680  
 30 TGGGACCTCG CGCGGGTTGA ACGATTTTAG TGTCCACAAC AGATTGAATC ATTTCTTTAT 4740  
 CCAAATTACG TGGTACATGA ACTTCCTTAG GTAAATATG TTGGTTTAAAG CTATAAAATT 4800  
 GTCCAATAAA TGTATAAAAT TCTTCTTCTT CTGTTTGCTG TAATGGAATC ATCGTTGTAT 4860  
 35 CTCGCTTTAT CATATTACCT TGTCGTATAA AGAAAACCTG GATACACATC CATCCTTTAT 4920  
 CAACACTATA ACCAAAGACA TCACGAATCG TTTTATCTGA TGACATAATT TTTTGTTTGT 4980  
 TTGTGAGATT TTGAATATGT TGAATTAAAT CTCTATATTC TTTAGCCCGT TCAAAATCAA 5040  
 40 GTGATTCACT TGCAGTTAAC ATTCGCTCTT CTAAACTTTT TAAAATTGTT TTGTCTTCCC 5100  
 CATTGAGAAA ATCAGTAATT TCCTTCGTCA TTTGTGCGTA TTTACTCAA TCAACGTCAT 5160  
 ATACACATGG TCCTAAACAT TGTCCAATAT GGTAATAAAG ACATAATTTA TCTGGCATCT 5220  
 45 TATCACATTT GCGATATGGA TATATTCTGT CTAATAACTT TTTAGTTTCT TGAGCAGAA 5280  
 ATGCATTCCG ATACGGTCCG AAATATTTGC CAGTACCTTG TTTTACAGTT CTCGTCACTA 5340  
 50 GTAGTCTAGG ATATTTCTCC TTCGTAATTT TAATAAATGG ATAACTTTFA TCATCCTTTA 5400  
 ATAATATATT ATATCTTGGT TGATATTGTT TAATCAGATT CAATTCCAGT AAAAGTGATT 5460

55

TTTTAGCATC ATGAGCACCC GTAAAATATG ATCGCAATC

5559

(2) INFORMATION FOR SEQ ID NO: 200:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4594 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

15 AAATCAATCG AGTGGCATGT CAAGGTCATA TCAATATTTT AGAATCTGCG ACTATGAGAG 60  
 AGGAAATAAA TGAAATTGCG CGACGTATCA TCGTTGATAT TCGTGATAAG CAATTACGAT 120  
 ATCAAGATAT TGCTATTTTA TATCGTGATG AATCTTATGC TTATTTATTT GATTCCATAT 180  
 20 TACCGCTTTA TAATATTCCT TATAATATTG ATACAAAGCG TTCGATGACA CATCATCCGG 240  
 TCATGGAAAT GATTCGTTCA TTGATTGAAG TTATTCAATC TAATTGGCAA GTGAATCCAA 300  
 TGCTACGCTT ATTGAAGACT GATGTGTAA CGGCATCATA TCTAAAAAGT GCATACTTAG 360  
 25 TTGATTTACT TGAAAATTTT GTACTGAAC GTGGTATATA CGGTAAACGT TGGTTAGATG 420  
 ATGAGCTATT TAATGTCGAA CATTTTAGCA AAATGGGGCG TAAAGCGCAT AACTGACCG 480  
 AAGATGAACG TAACACATTT GAACAAGTCG TTAAGTTAAA GAAAGATGTC ATTGATAAAA 540  
 30 TTTTACATTT TGAAAAGCAA ATGTCACAAG CGGAACTGT AAAAGATTTT GCAACTGCTT 600  
 TTTATGAAAG TATGGAATAT TTCGAACTGC CAAATCAATT GATGACAGAG CGAGATGAAC 660  
 TTGATTTAAA TGGTAATCAT GAAAAGGCGG AGGAAATTGA TCAAATATGG AATGGCTTAA 720  
 35 TTCAAATCCT TGATGACTTA GTTCTAGTAT TTGGAGATGA ACCAATGTCG ATGGAACGTT 780  
 TCTTAGAAGT ATTTGATATT GGTTTAGAAC AATTAGAATT TGTATGATT CCGCAAACAT 840  
 TGGACCAAGT AAGTATTGGT ACGATGGATT TGGCTAAAGT CGATAATAAG CAACATGTTT 900  
 40 ACTTAGTAGG TATGAATGAT GGAACGATGC CACAACCAAGT AATGCGTCAA GCTTGATTAC 960  
 AGATGAAGAA AAGAAATACT TTGAACAGCA GGCTAATGTC GAGTTAAGTC CAACATCAGA 1020  
 45 TATTTTACAG ATGGATGAAG CATTTGTTTG TTATGTTGCT ATGACTAGAG CTAAGGGAGA 1080  
 TGTTACATTT TCTTACAGTC TAATGGGATC AAGTGGTGAT GATAAGGAGA TCAGCCCATT 1140  
 TTTAAATCAA ATTCAATCAT TGTTCACCA ATTGGAAATT ACTAACATTC CTCAATACCA 1200  
 50 TGAAGTTAAC CCATTGTCAC TAATGCAACA TGCTAAGCAA ACCAAAATTA CATTATTTGA 1260  
 AGCATTGCGT GCTTGGTTAT ATGATGAAAT TGTGGCTGAT AGTTGGTTAG ATGCTTATCA 1320

55

EP 0 786 519 A2

	GTTTGACAAT GAAACTGTAA AATTAGGTGA AACGTTGTCT AAAGATTTAT ATGGTAAGGA	1440
	AATCAATGCC AGTGTATCCC GTTTTGAAGG TTATCAACAA TGCCCATTTA AACACTATGC	1500
5	GTCACATGGT CTGAAACTAA ATGAGCGAAC GAAGTATGAA CTTCAAACT TTGATTTAGG	1560
	TGATATTTTC CATTCTGTTT TAAAATATAT ATCTGAACGT ATTAATGGCG ATTTTAAACA	1620
	ATTAGACCTG AAAAAAATAA GACAATTAAC GAATGAAGCA TTGGAAGAAA TTTTACCTAA	1680
10	AGTTCAGTTT AATTTATTAA ATTCTTCAGC TTAATATCGT TATTTATCAA GACGCATTGG	1740
	CGCTATTGTA GAAACAACAC TAAGCGCATT AAAATATCAA GGCACGTATT CAAAGTTTAT	1800
	GCCAAAACAT TTTGAGACAA GTTTTAGAAG GAAACCAAGA ACAAATGACG AATTAATTGC	1860
15	ACAAACATTA ACGACAACCTC AAGGTATTCC AATTAATATT AGAGGGCAAA TTGACCGTAT	1920
	CGATACGTAT ACAAAGAATG ATACAAGTTT TGTTAATATC ATTGACTATA AATCCTCTGA	1980
	AGGTAGTGCG ACACTTGATT TAACGAAAGT ATATTATGGT ATGCAAATGC AAATGATGAC	2040
20	ATACATGGAT ATCGTTTTAC AAAATAAACA ACGCCTTGGA TTAACAGATA TTGTGAACCA	2100
	GGTGGaTTAT TATACTTCCA TGTACATGAA CCTAGAATTA AATTTAAATC ATGGTCTGAT	2160
25	ATTGATGAAG ATAAACTAGA ACAAGATTTA ATTAATAAAGT TTAAGTTGAG TGGTTTAGTT	2220
	AATGCAGACC AAACTGTTAT TGATGCATTG GATATTCGTT TAGAACCTAA ATTCACCTCA	2280
	GATATTGTAC CAGTTGGTTT GAATAAAGAT GGCTCTTTGA GTAAACGAGG CAGCCAAGTG	2340
30	GCAGATGAAG CAACGATTTA TAAATTCATC CAACATAACA AAGAGAATTT TATAGAAACA	2400
	GCTTCAAATA TTATGGATGG ACATACTGAA GTTGCAACCAT TAAAGTACAA ACAAAAATTG	2460
	CCATGTGCTT TTTGTAGTTA TCAATCGGTA TGTCACTAG ATGGCATGAT TGATAGTAAG	2520
35	CGATATCGAA CTGTAGATGA AACAATAAAT CCAATTGAAG CAATTCAAAA TATTAACATT	2580
	AATGATGAAT TTGGGGGTGA GCAATAGATG ACAATTCCAG AGAAACCACA AGGCGTGATT	2640
	TGGACTGACG CGCAATGGCA AAGTATTTAC GCAACTGGAC AAGATGTACT TGTTGCAGCC	2700
40	GCGGCAGGTT CAGGTAAAAC AGCTGTACTA GTTGAGCGTA TTATCCAAA GATTTTACGT	2760
	GATGGCATTG ATGTCGATCG ACTTTTAGTC GTAACGTTA CAACTTAAG CGCACGTGAA	2820
45	ATGAAGCATC GTGTAGACCA ACGTATTCAA GAGGCATCGA TTGCTGATCC TGCAAATGCA	2880
	CACTTGAAAA ACCAACGCAT CAAAATTCAT CAAGCACAAA TATCTACACT CCATAGTTTT	2940
	TGCTTGAAAT TAATTCAACA GCATTATGAT GTATTAAATA TTGACCCGAA CTTTAGAACA	3000
50	AGCAGTGAAG CTGAAAATAT TTTATTATTA GAACAAACGA TAGATGAGGT CATAGAACAA	3060
	CATTACGATA TCCTTGATCC TGCTTTTATT GAATTAACAG AGCAATTGTC TTCAGATAGA	3120

55

AATCTACAA ATTGGTTGGA TCAATTGGTG ACACCATACG AAGAAGAAGC ACAACAAGCG 3240  
 CAACTTATTC AACTACTAAC AGACTTATCT AAAGTATTTA TCACAGCTGC TTATGATGCT 3300  
 5 TTAATAAGG CGTATGATTT GTTTAGTATG ATGGATAGCG TCGATAAACA TTTAGCTGTT 3360  
 ATAGAAGATG AACGACGTTT AATGGGGCGT GTTTTAGAAG GTGGCTTTAT TGATATACCT 3420  
 TATTTAACTG GTCACGAATT TGGCGCGCGT TTGCCTAATG TAACAGCGAA AATTAAAGAA 3480  
 10 GCAATGAAA TGATGGTCGA TGCCTTAGAA GATGCTAAAC TTCAGTATAA AAAATATAAA 3540  
 TCATTAATTG ATAAAGTGAA GAGTGATTAC TTTTCAAGAG AAGCTGATGA TTTGAAAGCT 3600  
 GATATGCAAC AATTGGCGCC ACGAGTAAAG TACCTTGCGC GTATTGTGAA AGATGTTATG 3660  
 15 TCAGAATTCA ATCGAAAAA GCGTAGCAAA AATATTTTGG ATTTTCTGA TTATGAACAT 3720  
 TTTGCATTAC AAATTTTAAC TAATGAGGAT GGTTCGCCTT CAGAAATTGC CGAATCATAC 3780  
 CGTCAACACT TCCAAGAAAT ATTGGTCGAT GAGTATCAAG ATACGAACCG AGTTCAAGAG 3840  
 20 AAAATACTAT CTTGCATCAA AACGGGTGAT GAACATAATG GTAATTTATT TATGGTTGGA 3900  
 GATGTTAAGC AATCCATTTA TAAATTTAGA CAAGCTGATC CAAGTTTATT TATTGAAAAG 3960  
 25 TATCAACGCT TTAATATAGA TGGAGATGGC ACTGGACGTC GAATTGATTT GTCGCAAAAC 4020  
 TTCCGTCTC GAAAAGAAGT ACTGTCAACG ACTAACTATA TATTCAAACA TATGATGGAT 4080  
 GAACAAGTCG GTGAAGTAAA ATATGATGAA GCGGCACAGT TGTATTATGG TGCACCATAT 4140  
 30 GATGAATCGG ACCATCCaGT AACTTAAAA GTCCTTGTTG AAGCGGATCA AGAACATAGT 4200  
 GATTTAACTG GTAGTGAACA AGAAGCGCAT TTTATAGTAG AACAAAGTTAA AGATATCTTA 4260  
 GAACATCAAA AAGTTTATGA TATGAAAACA GGAAGCTATA GAAGTGCGAC ATACAAGGAT 4320  
 35 ATCGTTATTC TAGAACGCAG CTTTGGACAA GCTCGCAATT TACAACAAGC CTTTAAAAAT 4380  
 GAAGATATTC CATTCCATGT GAATAGTCGT GAAGGTTACT TTGAACAAAC AGAAGTCCGC 4440  
 TTAGTATTAT CATTTTTAAG AGCGATAGAT AATCCATTAC AAGATATTTA TTTAGTTGGG 4500  
 40 TTAATGCGCT CCGTTATATA TCAGTTCAAA GAAGACGAAT TAGCTCAAAT TAGAATATTG 4560  
 AGTCCAAATG ATGACTACTT CTATCAATCG ATTG 4594

(2) INFORMATION FOR SEQ ID NO: 201:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6313 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear



	GGTTTTCTnTG GAAAGATAGT GAAAATCTCG TGTTTTTTGG TTTTgAGGTG TTGTTTGTAT	60
	TTTaTAAaAT GGCTTACATA TATGAAGCGT TGATTAAGTA TGAaATTGTT AATTAATTGA	120
5	ACCTATTTAG CTTTAAGAAG GCATAACAAG ATGACCTTAT TTTATGCTAT AATATTTCTA	180
	TTATGCGAAG ATTAAGGTGA GTAGTAAATT GGATAAAAAA GTAAGTATTC AAACAAAGCA	240
	AGTGTTGAAA CAGCACAAAG AAAAAGAAAA ATTTGAATTT ACTACTGAAG GAACTTGGCA	300
10	ACAAAGGCAA TCTAACTTTA TTCGGTATGT AGAACAAATT GAGGATGCAA CAGTTAATGT	360
	TACAATAAAA GTGGATGATG ATAGCGTTAA GTTGATTTCGT AAAGGCGACA TTAATATGAA	420
15	TTTGCATTTT GTTGAAGGAC AAACGACAAC AACTTTTTTAC GATATATCGG CTGGACGAAT	480
	TCCACTAGAA GTTAAACAT TACGCATTTT ACATTTTCGT AGTGGAGACG GTGGCAAGCT	540
	AAAGATTCAT TATGAATTAT ATCAAGATAA TGAAAAAATG GGTCTTATC AATATGAAAT	600
20	TAActATAAG GAGATAGGCG AATGAATATT ATTGATCAAG TGAAACAAAC ATTAGTAGAA	660
	GAAATTGCAG CAAGTATTAA CAAAGCAGGA TTAGCAGATG AGATTCCTGA TATTAAaATT	720
	GAAGTTCCTA AAGATACAAA AAATGGAGAT TATGCTACTA ATATTGCGAT GGTACTGACT	780
25	AAGATTGCAA AGCGTAATCC TCGTGAAATT GCTCAAGCGA TTGTTGATAA CTTAGATACT	840
	GAAAAAGCAC ATGTAAaCA AATTGACATT GCTGGTCCAG GATTCATTAA TTTTACTTA	900
	GATAATCAGT ATTTAACAGC AATTATTCTT GAAGCAATTG AAAAAGGTGA TCAATTTGGA	960
30	CATGTAAATG AATCAAAAGG TCAAAATGTA TTGCTTGAGT ATGTTTCAGC TAACCCTACA	1020
	GGAGATTTAC ATATTGGTCA TGCTAGAAAT GCAGCAGTTG GTGATGCTTT AgcTAaATT	1080
	TTAACTGCAG CTGGCTATAA TGTAACACGT GAATATTATA TTAATGATGC TGGTAAATCAA	1140
35	ATTACTAACT TAGCGCGTTC GATTGAAACA CGTTTCTTTG AAGCTTTAGG TGACAATAGT	1200
	TATTCaATGC CAGAAGATGG CTATAATGGA AAAGATATTA TTGAAATAGG TAAAGATTTA	1260
40	GCAGAGAAAC ACCCTGAAAT TAAAGATTAT TCTGAAGAAG CACGTTTGAA AGAATTTAGA	1320
	AAATTAGGCG TAGAATACGA AATGGCTAAA TTGAAAAATG ATTTAGCAGA GTTCAATACG	1380
	CATTTTGATA ATTGGTTTAG TGAAaCATCT TTATATGAAA AAGGAGAAAT TCTTGAAGTT	1440
45	TTAGCAAAAA TGAAAGAATT AGGTTATACG TATGAAGCTG ATGGCGCTAC ATGGTTACGT	1500
	ACAACTGATT TTAAAGACGA CAAAGACAGA GTATTAATTA AAAATGACGG TACATATACG	1560
	TATTTCTTAC CAGATATTGC GTACCACTTC GATAAAGTAA AACGTGGTAA TGACATTTTA	1620
50	ATCGATTTAT TTGGTGCTGA TCATCATGGT TATATTAATC GTTTGAAAGC ATCTCTTGAA	1680
	ACGTTTGGTG TAGATAGTAA TCGTTTAGAA ATTCAAATCA TGCAaATGGT TCGTTTAATG	1740

55

	ATTATGGACG AaGTTGGCGT TGACGCTGCA CGTTATTTCT TAACTATGCG TagTCCTGAT	1860
	AGTCACTTTG ATTTTGATAT GGAATTAGCG AAAGAGCAAT CTCAAGACAA TCCAGTTTAC	1920
5	TATGCTCAAT ATGCACATGC GCGTATTTGT TCAATTTTAA AACAAGCGAA AGAGCAAGGT	1980
	ATTGAAGTGA CTGCTGCGAA TGATTTTACA ACGATTACTA ATGAAAAAGC GATTGAATTG	2040
	TTGAAAAAAG TAGCTGATTT CGAACCTACA ATTGAAAGTG CTGCTGAGCA TAGATCGGCA	2100
10	CATAGAATTA CTAATTATAT TCAAGATTTA GCTTCTCATT TCCATAAATT CTATAATGCT	2160
	GAAAAAGTGT TAACAGATGA TATTGAAAAA ACAAAGCAC ATGTTGCTAT GATTGAAGCG	2220
	GTCAGAATTA CATTGAAAAA TGCATTGGCA ATGGTCGGTG TAAGCGCACC TGAATCAATG	2280
15	TAAGAACATT TATATACACT CCAACGTAGA GTTTCTCGAA AGATACTTTG TGTGGAGTG	2340
	TTTTTTTAG GTATGTGACA TATTGGGGAA TGCTTAGTAT GTGAATAAGG TTAAGAGGAA	2400
	CACAGTTGGA TGCTCTGCAC AACTGCATAA GAGAGCCTGA GACATAAATC AATGTTCTAT	2460
20	GCTCTACAAA GTTATAATGG CAGTAGTTGA CTGAACGAAA ATTCGCTTGT AACAAGCTTT	2520
	TTTCAATTCT AGTCAACCTT GCCGGCGGGG CCCCACAAA GAGAAATTGG ATTCCCAATT	2580
25	TCTACAGACA ATGCAAGTTG GGGTGGGACG ACGAAATAAA TTTTACGATA ATATCATTTT	2640
	TGTCCCACTC CCTCTAAAAT GGAGGGTGTA AATGTTAGGA ACTGATGAAT TATATAAAGT	2700
	TTTATATGAA CATCTCGGAC CACAATTTTG GTGGCCTGCT GATAATGACA TTGAAATGAT	2760
30	GTTAGGTGCA ATTTTAGTTC AAAATACTAG ATGGCGAAAT GCAGAAATTG CATTGAATCA	2820
	GATTAAAGAA CATACGCATT TTAATCCAAA TCATATATTA GAACTACCTA TTGAAACGTT	2880
	ACAATCATTTG ATACATTCAA GTGGCTTTTA TAAAAGTAAA TCACTGACGA TTAAAACATT	2940
35	ATTAACATGG TTAGCACGAC ATCATTTCAA TTATCAAGAG ATTAATGAGC GATATAAAGG	3000
	TGGATTAAAGA AAAGAATTAT TATCTTTGAA AGGTATTGGA AGTGAAACAG CAGATGTCTT	3060
	ACTTGTTTAT ATATTGCGAC GTATTGAATT TATTCCAGAT AGCTATACAA GAAAAATATA	3120
40	TGATAAATTA GGATATGAAA AACTAAAAA TTATGATCAA TTAAAAAAG TAGTCAATT	3180
	ACCAAATCAT TTTACAAATC AAGATGCTAA TGAATTTTAT GCTCTGTTAG ATGTATTTGG	3240
45	TAAACATTAC TTTAGAGACA AAGATATAAA GAATTATGAT TTTTLAGAAC CTTACTTTAA	3300
	AAAGTAAACG CTGTGAAGTT AGATAGATGA GTTTATATGA AATATAAAAA ATAATTTACT	3360
	ATTTTCTTTT AGTATGTGGA CTTATATAAT AAATAGAAGC ATATAAAGAA AAAACAGTT	3420
50	GTTTGTGTGT GCAGCAACTG CATAAGAGCC CCTAATCGCT AAAGCTCAAG GGGAGTAAAG	3480
	GAATACAGTT GTTTGTGCAG CAACTGCATA AAAGCCTCTA ATCACTAAAG GTGAAGAGGA	3540

55

## EP 0 786 519 A2

	AACGCAGTTG	GATgCTACCG	CACAACTGCA	TAAATCCCTC	TaATCgCTAA	AGCGAAAAGT	3660
	GGGATTAAAA	AGGAGATGTG	ATAGTGTGAA	GAAATCGTTA	ATTGCTTTTA	TTTTGATTTT	3720
5	TATGCTTGTC	CTGAGTGGCT	GTGGTATGAA	AGATAATGAT	AAACAAGGTA	GCAATGATAA	3780
	TGGCTCGTCT	AAATCGCCGT	ACCATAGAAT	TGTTTCGTTA	ATGCCTAGTA	ATACTGAAAT	3840
	TTTATATGAA	TTAGGATTAG	GTAAATACAT	AGTTGGTGTT	TCAACGGTTG	ATGATTATCC	3900
10	AAAAGATGTG	AAAAAGGGTA	AGAAACAATT	TGATGCTTTG	AATCTAAATA	AAGAGGAACT	3960
	TTTAAAGGCA	AAGCCAGATC	TAATTCTTGC	GCATGAGTCG	CAAAAGGCAA	CTGCTAATAA	4020
	AGTATTGTCA	TCATTAGAGA	AACAAGGCAT	CAAAGTAGTG	TATGTTAAAG	ATGCACAATC	4080
15	AATTGATGAA	ACTTACAACA	CATTTAAGCA	AATTGGGAAA	TTAACGCATC	ATGATAAGCA	4140
	GGCTGAACAA	CTTGTTGAGG	AACTAAAGA	TAATATCGAT	AAAGTCATAG	ATTCAATTCC	4200
20	TGCTCATCAT	AAAAAATCAA	AAGTATTTAT	TGAGGTTTCA	TCAAAGCCTG	AAATATATAC	4260
	AGCAGGGAAG	CATACATTTT	TTAATGATAT	GTTAGAAAAA	TTAGAAGCCC	AAAATGTGTA	4320
	TAGTGACATT	AATGGTTGGA	ACCCTGTAAC	GAAGGAAAGT	ATTATTAAAA	AGAACCCAGA	4380
25	TATATTAATT	TCGACGGAAG	CTAAGACAAG	ATCAGATTAT	ATGGATATCA	TCAAAAAAAG	4440
	AGGTGGATTG	AATAAAATTA	ATGCTGTCAA	GAATACACGT	ATTGAAGTTG	TAAATGGTGA	4500
	TGAAGTATCA	AGACCAGGTC	CACGTATTGA	TGAAGGATTA	AAAGAATTAA	GAGATGCAAT	4560
30	TTATAGAAAA	TAAACCATTG	TAATTATGCC	CCTTATTGCT	ACATGTAAAA	AATACATGTT	4620
	TGAGATAAGG	GGTTTTTaAA	ATATATTTAG	TGAATGATAG	CAACGCGAGT	ATGTGATTGC	4680
	TATAATGAAT	GTAATTATCG	ATGAaCaAAA	GAGAATGCTA	TGACATTTAA	TAAAGTATTA	4740
35	TTGAGCTGga	TAGTCmTATT	GATTATAACA	ACTAGCATAT	ATCTATTTTG	GCAGTTGGGC	4800
	GATATCAATG	ATGTATTTAA	CCAGTCTATT	TTAATCAATG	TTAGATTACC	GAGATTATTA	4860
	GAAGCATTGT	TGACAGGTAT	GATATTAACT	GTTGCAGGCC	TTATATTTCA	AACAGTTTTA	4920
40	AATAATGCAT	TGGCAGATAG	CTTTACATTA	GGATTGGCAA	GCGGCGCTAC	ATTTGGTTCA	4980
	GGATTAGCAT	TATTTTtagg	TTAACAACG	TTATGGATTG	CTGTATTTTC	AATAACATTT	5040
45	AGTTTGATAA	CATTAATAAC	TGTATTAGTC	ATTACGTCGG	TATTGAGCCA	AGGCTATCCA	5100
	GTTAGAATCT	TAATATTAAG	TGGTTTAATG	ATTGGTGCGT	TATTCaATTG	ACTTCTATAT	5160
	TTTTTGATTT	TATTAAAACC	TCGCAAATTA	AATACAATTG	CCAATTATCT	GTTTGGTGGT	5220
50	TTTGGTGATG	CAGAATACTC	AAATGTATCT	ATAATAGCAA	TCACATTTAT	CATTGCATTG	5280
	TTTGGTATAT	TTATCATTCT	TAATCAACTA	AAGTTATTGC	AATTAGGAGA	ACTAAAAAGT	5340
55							

EP 0 786 519 A2

ATAACGGCGA TAAATGTGCG ATATGTTGGC ATCATTGGAT TCATTGGTAT GGTGATACCG 5460  
 CAACTCATTG GAAAATGGCA GTGGAAACAA TCATTAGGAA GACAATTGGC TTAAATATT 5520  
 5 GTAAGTGGAG GACAAATAAT GGTATGGCA GATTTTATTG GTAGCCATAT ATTGTCACCA 5580  
 GTACAAATAC CGGCAAGTAT TATCATTGCA TTAATTGGTA TACCAGTGTT AtTTTACaTG 5640  
 CkAAwAtCtC aGTCgAAAcG GTTACaCTAG CACACGACaT TTGCTAAAAT AAAAATAACT 5700  
 10 ATAAACATAA AGAGGGCATA AGCGATGGAT TTGAATCAAA TTAAAGCAGT TGTATTTGAT 5760  
 TTAGAAGGTA CGTTGTTGGA CAGAGTTAAA TCTCGAGAGA AATTTATCGA AGAGCAATAT 5820  
 GAACGATTTT ATGACTACTT AATTCATGTT CAACTGGCAG ATTTTAAAAA AgCATTATT 5880  
 15 GAGCTAGATG ACGATGAAGA TAATGATAAA CCTGATTTAT ATAAAGAAAT CATTAAACGT 5940  
 TTCCATGTAG ATAGGTTAAC TTGGAAAGAC TTATTTAATG ATTTTGAAAT GCATTTTAT 6000  
 CGTTATGTAT TTCCTTATTA CGATACTTTG TATACACTAG AAAAgCTATC GCAAAAAGGC 6060  
 20 TTTCAAATTG GTGTTATCGC AAATGGTAAA TCTAAGATTA AACAAATTCG ATTACATTCA 6120  
 CTGGGTTTGA TGCATGTTAT TAATTATTTA TCAACATCAG AAACAGTTGG TTTTCGTAAA 6180  
 25 CCACATCCTA AAATTTTGA AGATATGATT GATCAACTAG GGTATTACC TGAGCAAATT 6240  
 ATGTATGTTG GCGATGATGC GTTAAATGAT GTAGCTCCAG CACGAGCTAT GGGCATGGTT 6300  
 AGTGATGGT ATA 6313

30 (2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2174 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

40 CCGTAAACAC ATCAACAAAA GAAGGCTATA TTACAAAAGA AGACTTGGAC TTATGCTGCA 60  
 CGTCGCTCTA ATTCAGCTGG AATGCAAGTC ACCGGACGAC TGGCTTACAT TGAACCTTAT 120  
 GGGGCAACAA GTCGCACAAA ATAAACGCGC GAGAAGCaAG AATAGGAAGT GATATCTATG 180  
 45 AAATGGTTAT CACGAATATT AACAGTAATA GTGACCATGT CtATGGcGTG TGGTGCAATTG 240  
 ATATTTAATC GTAGACATCA GCTAAAGGCG AAAACGCTGA ACTTCAATCA TAAAGCATTA 300  
 50 ACAATTATTA TTCCGGCTAG AAACGAAGAA AAAAGAATAG GTCATTTACT ACATTCGATA 360  
 ATACAACAGC AAGTTCCAGT AGATGTCATT GTTATGAATG ACGGATCGAC AGATGAAACA 420

55

	AAATGGTATG GGAAATCACA TGCTTGTTAT CAAGGTGTGA CGCATGCATG TACGAATCGC	540
	ATTGCCTTTG TAGATGCTGA TGTAACTTTC TTAAGGAAAG ATGCTGTTGA AACGTTGATT	600
5	AATCAGTATC AATTACAAGG TGAAAAAGGA TTGTTAAGCG TACAGCCTTA TCATATAACA	660
	AAGCGTTTCT ACGAAGGGTT TTCAGCGATA TTTAATTTAA TGACAGTCGT TGGTATGAAT	720
	GTATTTTCTA CCTTAGACGA CGGTCGACT AACCAGCATG CATTTGGACC GGTGACATTA	780
10	ACAAATAAAG AAGATTATTA TGCAACTGGA GGTCAATAAA GTGCAAACCG TCATATTATT	840
	GAAGGATTTG CTTTAGGAAG TGCATATACT TCACAATCAT TGCCCGTAAC AGTTTATGAA	900
15	GGGTTTCCAT TTGTTGCATT TCGCATGTAT CAAGAAGGAT TTCAGTCATT ACAAGAAGGA	960
	TGGACAAAGC ATTTGTCAAC TGGGGCAGGT GGCACAAAGC CTAAGATCAT GACAGCAATT	1020
	GTGTTGTGGT TGTTTGGTTC TATAGCGAGT ATTTTAGGGC TATGTCTTAG TTTAAAATAT	1080
20	CGCCAAATGT CTGTAAGAAA AATGGTAGCA CTTTACTTGA GCTATACTAC ACAATTTATT	1140
	TATCTGCATC GAAGGGTCGG CCAATTTTCT AATTTATTAA TGGTATGTCA TCCATTGTTA	1200
	TTTATGTTTT TTAATAAAAT TTTCATCCAA TCTTGAAAC AAACGCATCG TTATGGTGTA	1260
25	GTTGAATGGA AAGGTCGTCA ATATTCTATA TCTAAAGAAC AATAAATCAA GGTAATGGCA	1320
	TTTCAATATA GGAGGACTAG TATGACAATG ATGGATATGA ATTTTAAATA TTGTCATAAA	1380
	ATCATGAAGA AACATTCAAA AAGCTTTTCT TACGCTTTTG ACTTGTTACC AGAAGATCAA	1440
30	AGAAAAGCGG TTTGGGCAAT TTATGCTGTG TGTCGTAAAA TTGATGACAG TATAGATGTT	1500
	TATGGCGATA TTCAATTTTT AAATCAAATA AAAGAAGATA TACAATCTAT TGAAAAATAC	1560
	CCATATGAAC ATCATCACTT TCAAAGTGAT CGTAGAATCA TGATGGCGCT TCAGCATGTT	1620
35	GCACAACATA AAAATATCGC CTTTCAATCT TTTTATAATC TCATTGATAC TGTATATAAA	1680
	GATCAACATT TTACAATGTT TGAACCGAC GCTGAATTAT TCGGATATTG TTATGGTGTT	1740
40	GCTGGTACAg TAGGTGAAGT ATTGACGCCG ATTTTAAGTG ATCATGAAAC ACATCAGACA	1800
	TACGATGTCG CAAGAAGACT TGGTGAATCG TTGCAATTGA TTAATATATT AAGAGATGTC	1860
	GGTGAAGATT TTGACAATGA ACGGATATAT TTTAGTAAGC AACGATTAAA GCAATATGAA	1920
45	GTTGATATTG CTGAAGTGTA CCAAATGGT GTTAATAATC ATTATATTGA CTTATGGGAA	1980
	TATTATGCAG CTATCGCAGA AAAAGATTTT CAAGATGTTA TGGATCAAAT CAAAGTATTT	2040
	AGTATTGAAG CACAACCAAT CATAGAATTA GCAGCACGTA TATATATTGA AATACTGGAC	2100
50	GAaGTGAGaC AGGCTAACTA TACATTACAT GAACGTGTTT TTGTGGaTAA GAGGAAAAAG	2160
	GCAAAGTTGT TTCA	2174

55

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4715 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

10 GAAnCAGnTA GACAAATTAT GGaAamCGGT GTGAATCaAG GATTcTTGG TGTAGCTGGT 60  
 TTTGACCTAC TCGTCGATGA GGATGATAAC GTTTATGCGA TTGATTtAAA CTTTAGACAA 120  
 AATGGTTCaA CGAGCATGTT ATTACTTGCT AACGAGTTGA ATTcAGGATA TCAAAAGTTT 180  
 15 TATAGTTATC ATTCAAAAGG TGATAACACA CATTTCTTCA ATACGATTTT GAAATATGTC 240  
 AAAGAAGGTA GTTTATACCC GTTATCTTAT TATGATGGTG ATTGGTACGG TGAAGATAAA 300  
 GTTAAATCAA GGTtTGGCTG TATTTGGCAT GGTGATTCAA AAGAAACAGT ACTGGAGAAT 360  
 20 GAACGCGCAT TTTTAGCTGA ACTTGAACAC TATTAGAGTT CGGAACATAA GGCGCTACAA 420  
 TGTtGTGTG CCAGTAGTTG ACTGAATATG CGTTTGTAAc AAGCTTTTTT CGATTCTAGT 480  
 25 CAACAGTAAT TAAATTTATG ATATGGCAAT ACTTTGTAAT ACTAATATTA AATGGCGACT 540  
 TTTATTTcAC TATGTTATAA GAGTTGCCAT TTTGTGATA AAGGTATACT AAAGGTTATC 600  
 GTTTTGAAAT TTTTAGTAAC TAGATATGTT TCGTGTtATA GACCGAATTT GTGTATACGT 660  
 30 AAAATTTAAT GCTATTGAAT TTTTAAATG AAAAACATGA CATTAaATTG AATTCATAAT 720  
 ATGTCTAATT GACTAACTTG TTGGAGTCAT TtACTATTTT ATGTATGACA TATTTTAAAA 780  
 AGTGAGGGTC AAGCATGTCT TATAAAGCAT ATCCATtCTT TAGAGATATA TTAATAAATG 840  
 35 AATGTATTTA TTTCGCCTCT AAAAATAAAA AACTAGTACG CCTAAATTAT AAAAGTGAAG 900  
 CGnATGTAGG CGTTTGGACA GAAGAAAGTG TGGCCGTATC ATTTTTAACA AGTCGTGATA 960  
 TTCCATTTGA TAAAGTTGTA AAAATGGACG TTGATCGTTT TGCTACTTAT GAATTAGATG 1020  
 40 AATTGTTTGA TGAACAAGAC CATATTATTA TGAATCAAAC AATGGAAGAw GAAGGGCATC 1080  
 TACTAAACGT TGTAGCTGTT ACACAAGAAG TGATGACGGA ATTAGATAAA ATTAGAATCA 1140  
 AAGAATTTGT CCAAGATGTA GCGAAATATG ATGAAGTATA CGGCTTAACT AAAAAAGGTA 1200  
 45 GTAAGCAGTT TATTCTCATT AGTGAaaATG ATAGCGACGA AAAAAAGCCG CATATTATGC 1260  
 CTGTATGGAG TATTAaaaaAC AGAGCGTTAA AAGTTcGAGA TGAAGATTTT GAAGAGTGTG 1320  
 50 ATTTAATTAC GATTGAAGGT TCTGTTTTTCG GAGAATGGCT AGATGAaCTT AGAGATGATC 1380  
 ATAAAGCCGT TGCGATAGAT TTAAAACTG GCGTGGTTGG TACAATTGTT TCAGCGCAAA 1440

55

EP 0 786 519 A2

	ATGGAACAAT	ACGTATTCAA	AACACTTAGA	CCATAAAATA	AAAGGCCATT	TATATAGCGT	1560
	TTATTTAAAA	CAACGCGCAT	ATAAATGGTC	TTTTTCTATT	TTTCTAAATA	TAATGCACCA	1620
5	ATAGCACCTG	AAAAATGCGC	CGTTTTCAAC	ATAGTACGGT	TTGCAACCGC	GTAACACAGT	1680
	ATAATCTTCC	ACAACCTTGC	GTAATAAAGC	GTTATTATGA	AATGAAGAAC	CGATATAAAC	1740
	GATATTTTCA	GTTTTAAATT	CACGTGCAAC	AGTAATGGCC	ATTGTCGTAA	CAACTTCGCC	1800
10	AACGACACCA	ATAACGGCTG	CTAATTTATT	GCTAGGTGTA	AAATCAGCAT	CTAAATGATG	1860
	TAGTACATGA	CCAAAATTAG	CTGCTGTAA	ATCACCGGGA	ATGGGTGGTT	CGGTATCTTT	1920
15	ATAAATATGT	CTAACCTTTA	AATCGATAGT	GTTACGATCA	CCGTGTTGTG	CCATGTCAGT	1980
	TAAGTGTTA	TAATCAGTGA	TTTGACTTAG	TAAATAACCG	AGTCCTTGAA	TCATGCCTCC	2040
	ACCTGTACCG	ATACCGCCTA	CACGACGTTG	TGATTGGCCG	TCGAAATAAT	GTAGTGACGT	2100
20	ACCGGTACCA	ACATTTGCAA	AAATATAATC	TGCTAAGTCA	TGGCCTTGCT	CTTTTAACAA	2160
	AATACCTAGT	CCTTGAGATG	CAGCATCAAA	CTCTACAAA	ATTTGTGCAG	GAATGTTGAT	2220
	GTTTTTCAGCA	ATGACACCTG	CATTACCTCC	AGTTAAGCAT	AATTTTTCAA	TTTGCTGTTG	2280
25	GTTTAACCAT	TCCACAACCT	GATCAATATT	TTTAGTTAAT	TCAGTTTAA	AAGTACGTTG	2340
	GTTATCTTGC	TCTTGAACGA	TTTAAATTAG	TGTACCGCCA	GCGTCAATGC	CAACTTTCAT	2400
	AAGATTCCCA	CCTCATTATT	AATGTCTATC	CTTAAATAAT	AGTATAGTAA	AATGACTAAA	2460
30	AAACAAGTAA	TAATAGTAAT	TATTAACAAA	TTTGATGCCA	TTGCATTTC	ACATTGTAAG	2520
	CGTATCGCAA	TTAACTGTTT	ACAAACGTGG	ACGTTAAGTT	ATATATATTA	TTTTCTAGGA	2580
	ATTTTGAAGT	TGTATAGGAT	TGTTAGTTAG	TGACGCAATA	TTAAAAGTAG	TTCGTACGCA	2640
35	GTGTATTTGT	AAGTCTCTGA	TTAAAATGAT	AAGTAATGAG	GAATAGTACA	TTAATTTTGA	2700
	AATTTAAAAA	ATATAAATAA	GTAATTTATT	TAAGTTAGAG	CAAATAATGG	TATCGTAGTG	2760
40	AAATAATAGG	TAAAATAATA	TGGGGATTCA	TGCTTCATAT	ATAAAAAGAT	AGGGGTAA	2820
	TATATGGCTA	AAGAACTTTG	TTTGAAGGT	ATCACTTTAA	AAGCATTTGA	TGAACAATAT	2880
	CGTTCAGCAA	TTAATGATTT	TGACTTGAAT	GAAAGACAAC	AAATATATTC	ATCTTTACCT	2940
45	AAAGAAGTTA	TTGATGATGC	AATTAATGAT	GCTGATAGGA	TTGCTAACGT	AGCAWTAAMC	3000
	GATAAAAATG	AAGTGGTGGG	CTTTTTTGTA	TTACATCGTT	ACTATCAGCA	TGAAGGTTAT	3060
	GATACACCTG	AAAATGTCGT	TTATATTCGT	TCATTATCGA	TTAATGAAAA	ATATCAAGGT	3120
50	TTTGGATATG	GCACGAAAAT	AATGATGTCA	TGCGCGCAAT	ATGTTCAAGG	TGTATTTCCCT	3180
	GATTTTAATC	ATCTATATCT	AGTAGTAGAT	GCGGAAAATG	ACAATGCTTG	GAACCTATAC	3240
55							

CTATATTACT TGGACTTAGA TTCAAAACAT GTTTCATCAT TAAAGCTTGA AGAAGAAAAGT 3360  
 CGTTCAGAAG TGACCAATGT ACATATCATT AATTTAATGA TTGATGGCCA AAAGGTTGGC 3420  
 5 TTTATCGCAT TGGAGCAGAT TGGTGAACGC ATGAACATG CTGCTATTGA AGTGGATAAA 3480  
 TCATATCGCT TTAATGGTAT TGGTTCAAGT GCTCTGCGAC AATTGCCAAC TTACTTAAGA 3540  
 AAAAActATG ACAACCTTAA TGTGATTACG ATGATTCTGT TTGGAGAGAA TAATGATTTT 3600  
 10 AAACCATTAT GTTTAAATAG TAATTTCTGT GAAATCGAAC AACTGATGA TTATGTCGTT 3660  
 TTCGAAAAT ATTTAAATTA CTAACAGTGA TTGCGAAATA TGATATTGTC ATTTATAATT 3720  
 15 TAGTTTTGTT ACTATATATA AATGAATTCA GACGTATAAA TTTAGATTAT ATCCTTCGAA 3780  
 AGGAAGTATT GGGCAATGAA AATTCAAGAT TATACAAAAC AAATGGTTGA TGAAAAATCA 3840  
 TTTATTGATA TGGCTTATAC ATTATTGAAT GATAAAGGCG AAACAATGAm mTTATATGAT 3900  
 20 ATyATCGATG AATTTAGAGC GTTAGGTGAT TATGAGTACG AAGAAATTGA AAATCGTGTT 3960  
 GTACAATTTT ACACGGATTT AAACACAGAT GGTCTTTTTT TAAATGTTGG AGAAAAATTA 4020  
 TGGGGATTAC GTGATTGGTA TTCGGTAGAT GATATTGAAG AGAAAATCGC ACCAACTATT 4080  
 25 CAAAAATTCG ATATTCTGGA TGCAGATGAT GAAGAAGATC AAAACTTAAA ATTATTGGGC 4140  
 GAAGATGAAA TGGATGACGA CGATGATATT CCAGCTCAAA CAGATGATCA AGAAGAACTA 4200  
 AATGATCCAG AAGATGAGCA GGTGAAGAA GAAATCAATC ATTCGGATAT AGTCATTGAA 4260  
 30 GAAGATGAAG ATGAAGTAGA CGAAGACGAA GAAGTGTTTG AAGACGAAGA AGACTTCAAC 4320  
 GATTAATTTT TTGTTTGACT TTTAGTTGAA AGATGATAAA ATTTTATTCG GGCTCCTTTA 4380  
 AATAGGACAC GTGTATAAAA TTTATACGCT CCCCTTACAG AATTTGTGAG AGGGAGCGTT 4440  
 35 TTTTtATTTA ATTGAGTAAA TCAAGAAATG ATAACGCAAA AATCAAAGTT GTAAATGATA 4500  
 TACATAGTGA CATAGCAGTA TGGAAACGGT AAGTAAACAG AATTTAATTT TGTCGAtTCG 4560  
 40 ACAAtAAaCA aCTtGAaTGA GCTTGCTTTA ATGTTATGTn nTACGTAATT TTTACAATTG 4620  
 ATGAGGAAGC ATTCCCTTTA ATAATTAGGA GGTCAAGACA TGACAAAATT TATTTTGTGA 4680  
 ACAGGTGGCG TAGTTTCATC CATTAGGGGA AGGGT 4715

45 (2) INFORMATION FOR SEQ ID NO: 204:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 918 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

55



ATAATAACTG AAATTAAAAT TGCTAAATmG TGTtaAgCTA TCGCmACAAT GAAAATwCCG 60  
 ATTTTGCGTT GTTGAAAATA TCTTTCCAAA CCAAGAATCG ATAATGGCAA TAAATATAAT 120  
 5 AAATTTCCAT AAAATGACCA AGTAAAATTA AAGTATATAA CGACAGTTGA CATGCCGTAT 180  
 AAAATCGTAG CGATCATATT TGCTGAGCGT TTAAAGTGTA ATATTTTAAA TAAGTAGAAG 240  
 10 GTCACGACAA ATGTTATGAT AGCTCGTATC ATAGCCATAA TAAGTTGGTT TGTCGGCCAA 300  
 AAATGTATTG TCGTCGGATT AAATATACCA ACCGTTTCTC CTATTTTAAT GAAKAGAAAA 360  
 TTTAGCCACA TTAAAGGTGA CAGCGAATAA TAATnTGATA GTCCTTTTCAT ATAATCGCCA 420  
 15 CCTAmTCCAA ACGATGCATC ATrTAAACTA GAAaAACTAC GTAGATGTTT ATACAnATAC 480  
 ATTTGAAATG GCATCATTTG ACGGAATCCA TCTCCAGCCC CGCTAAAAAC AGTACCATTG 540  
 ACAATATAAT CATAGATATG AGTAGAAAAT AAAATAAGCG TTAATATTAC ACTAATGAAA 600  
 20 GTTATAACAA AGAATTGTTT GACGTTTGAA TTTAGCCACT TTTTAAACAC AACATTATCC 660  
 TCAACTTTCA AATTTAAAAT TAAGTTTAAAC TGAAACTAAA GTTAATGAGG TTCTTGATAG 720  
 GTAAAGACGA AGATGACTGT GGAACAGATA CCTTATCATA GTTACTTAAA CTTTGGATCA 780  
 25 TTTTCAGTTT ATCATTAAC AAATATATTG AATAATAAAa aTGTCATACT GATAAAGATG 840  
 AATGTCACTT AATAAGTAAC TTAGaTTTAA CAAATGATGA TTTTAAATTG TAGAAAACCT 900  
 GAAATAATCA CkTATACC 918

(2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16397 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

TCGCCChATA ATCAATTTAT TTTTCATGTG CCACTCCTAT ACAAGCTnAC AATGCTTCTT 60  
 CAGTTAAGGC AATATCTTTT AATTTTGTTT GATATTTTTG TTCAAAGTCA TATTGTAACT 120  
 45 GAACAATTTT TGGCAAACCA ATATGCCAAT CGCCAATTT TTTTTTAyCT TtGAAGAGCT 180  
 CTTTTGGTGA TGkTTGcGAC ACTATACTAC CTTCCTTCAT AACGATGACT TCATCTGCAT 240  
 AACGCGCGAC TTCATTCATA TCATGTGAAA TTAGGATAAT TGCCTTATTT TCATCTGTTT 300  
 50 GTAGTGACTT TAGTAATCTC ATTACTTGTC GTTACTTTG TGGATCAAGT CCTGCTGTAG 360  
 GTTCATCAAC CACGATAATA TCAGGATTCA TTGCCAATAT CGATACAATC GCTATTTTAC 420

	AATCCATCAA CAGACGATGG GCATAGTTTT TGGCTTCATC TAAATTCATT TTAAAGTTTT	540
	TAGGTCCAAA TATCATTTCA CGCTCTACTG TGTCCTCAAA TAATTGAGAT TCGGGAAATT	600
5	GAAATACCAT TCCAATTCTT TTTCTTACAG GTCTAATATA TTTATCTTTG GTCTTATGTG	660
	TAATAGTAAT GTCATCAACT GTAACGTGCC CAGTAGTCGG CTTTAACAGC GCATTAATAT	720
	TTTGATCAA CGTTGATTTA CCACTACCCG TTTGTCCAAC GATGGCGTAA TATTTACCTT	780
10	GTTCAAATTC TGTATTAAAC TCATGAATAG CTTGATGCTG ATATGGTGTC CCTTTTTGAT	840
	AGGTATAACT TACATTGTCA AACCGTATAG TCATAGTTGA TCCACCAGCC CTTCATAAGT	900
15	TAAGAATGAT GTTTGGTGTC CCAGCATTTG ATTTATTTTG ATTGGGAATG GCAAATCTAG	960
	ACCTATTCTT GTTAACTCTT CTGCATTGTC GAAAATTTCG GTCGCTGTGC CTTCTTTATA	1020
	GACAGTCCCT TTATTCATAA CGATAACATG ATCTGCTTCC ATCGCCTCAG ATAAATCATG	1080
20	CGTAATAGAA ATGATTGTAA TATTATGTTC TGATTTAACT TTTCTCACTA AATCCAATAA	1140
	ATTTTGACGT GCATCAGGAT CTAACATAGA AGTCGCCTCA TCTAATATAA TGACAGAGGG	1200
	GTAAAGTGCT AATACACTTG CTATAGCCAC ACGCTGCTTC TGTCCCCCG ATAATGCATT	1260
25	AGGTTTATAA TCTGCACGTT CTAACATATC AACTTGTTTA AGTGCTTCGC TGACTCTTCT	1320
	ATGCATTTTC TCATATGGAA CCGCATGATT TTCGAGTCCA AATGCCACAT CGTATTTTAC	1380
	AATTGAACCA ACAAATTGAT TATCCGGATT CTGAAATACA ATTCCTATGT CTTTCTTAA	1440
30	CTTTTCAAAA TTATCATCAG TTATAGCTTG ATTATTATAA AAAATTTCTC CAGATTTAAC	1500
	TTTCTCTATG CCAATCATT ACTTGGCAAT TGATAGTTTT CCAGAACCGT TATGACCAAC	1560
	AATAGATGTC CACTGACCTT TAGGTATATT AAAAGAAACA TCTTTCAATG TGAAGGATGC	1620
35	ATCACTTTGA TATTGAAATG AAACATTTTT AAATACAATA ACTGAATTCT TATCCTCCAC	1680
	TTGTCTCTCT CCTTTACGAT TCGTGTATCT ATCATATTTT ACAATATTTA TAAATCGCTG	1740
	TATATGACAT TGACTGGGTT CTCTATATAT TACTAGTATT TTCTGACTCA TTTCTAGTCT	1800
40	TTAAAGTGTT GTTTAACAAC TAATGATAAG GACTTTTATT CCTCTCTAAC AATTATGTAT	1860
	AAACGTTAAT AAAATAAATG ATTTACTAAT ATAGGGGTGG TCGCGTTTGA TTCAACGATA	1920
45	ATACTTTCAC TTCATTCACT TCTAGTGAAA TTGATCAAAC TAGCTTCATC ATATTTTTAG	1980
	ATTCGCACTC AAAAAAGTAA ATATAAAGAA ATCGGACTTA AAAACATTTT TGTTCATAAG	2040
	TCCGATATTT TATTCAATAA AAAAGCGCGC ACCCCATCAT AAGTTTGTTG AGTTCACGCT	2100
50	TTAAATCTTT ATTTAGTTGA TGGGGTACTC TGAGCTAGAC AATATTTGTA TGTGGCAAAC	2160
	ATTATCGTTG CACTCATTG CTTTATATAA AAGTAGTTAG TGTATTTATA TAAATCTTAA	2220

55

ACGAGTGTA CCACCTTGAC GTTCTGTGTA AcGCTCTGCG ATTTACACCA ATAATTTTGT 2340  
 AAGTGCAGTT TGTGTAGTTT CATCTTCGTT TAAGATTTCA ACATTACGTA AAGTTTTAGC 2400  
 5 TGCATTACGA CGAGAAGCTA AATCTCCTTT TTTACCTAAA GTGATTAATT TCTCAACAAC 2460  
 ACTGCGAACT TCTTTTGCAC GAGCTTCTGT AGTTTCAATA CGTTCACTAA TAATAAGTGA 2520  
 TGTAGCTAAG TCACGTAACA TAGCTTTACG TTGATCAGAA GTACGACCTA ATTTTCTGTA 2580  
 10 ACCCATGAGT TAACCTCCTT TATCAATCTT CTTTCTTAA TCCTAATCCT AAATCTTCTA 2640  
 ATTTGTATTT AACTTCTTCT AAAGATTAC GACCTAAATT ACGCACTTC ATCATGTCAG 2700  
 CTTCAGATTT GTCAGCTAAC TCTTGAACAG AATTGATTCC TGCGCGTTTT AAGCAGTTAT 2760  
 15 ATGAACGTAC AGATAAGTCT AATTCTTCAA TAGACATTTT TAATACTTTT TCTTTTTGAT 2820  
 CTTCTTCTTT TTCAATCATG ATTTCAGCGT TTTGCGCTTC ATCAGTAAGA CCAACGAAGA 2880  
 TATTCAAGTG TTCAGTCATT ATTTTGTCTG CTAATGAAAC TGATTCTTGT GGTGTGATTG 2940  
 20 AACCATTAGT CCAAACATCC AATGTTAATT TATCAAAATC ACTGCTTTGA CCTACACGTG 3000  
 TATTTTCAAC AGTATAGTTC ACACGTTCAA CAGGTGAATA CAATGAATCA ACAGGGATTA 3060  
 25 CACCAATTGG TAAATCACTA GTATTATTTT GTTCTGCTAA TGCGTAACCT CTACCCTTGT 3120  
 TAGCAACTAG ACGAATTTTT AAGTGACCAC CTTTAGATAC TGTGCAATT TTAAGCTCTG 3180  
 GGTTTAAAAT TTCAACATCA CTATCATGTG TAATGTCGCT TGCTGTTACT TCGCCTTCAT 3240  
 30 CACGTACATC AATTTCTAAA GTTTTATCTT CTTCAGAGTA AATTTTCAAT GCTAATTGTT 3300  
 TAATGTTTAT AATAATTGTA GAAACATCTT CAACTACATT GTCTACTGCT GAGAATTCAT 3360  
 GTAAACTCC CTCAATTTCA ATATACTTAA cGGCTGCACC TGGTAATGAA GATAGTAGGA 3420  
 35 TACGACGTAA GGAGTTTCCT AGTGTAGTAC CGTAGCCACG TTCTAGTGGT TCAACAACGA 3480  
 ACTTACCGAA TTTAGCATCT TCACTAATTT CAATTGTCTC AATTCTAGGT TTTTCGATTT 3540  
 CTATCATTTA AATATCCTCC TTATATACGT CGACTTAATT TAAAATGTTT GCTCAGTGAC 3600  
 40 CTGTAACAAT ACCATCATAA ATTATACACG ACGACGTTTT GGTGGACGAC AACCGTTATG 3660  
 AGGTAAGTGA GTAACGTCTC TGATCGCAGT TACTTCTAAA CCTGCAGATT GTAATGCACG 3720  
 AATAGCTGAT TCACGACCTG GACCAGGTCC TTTAACTGTT ACTTCAACTG TTTTAAACC 3780  
 45 ATGCTCCATA GCTGATTTAG ATGCAGTTTC AGAAGCCATT TGTGCTGCAA ATGGTGTGTA 3840  
 TTTTGTAGAT CCTTTGAATC CTAATGCACC AGCTGATGAC CATGATAAAG CATTACCGAA 3900  
 50 CTCATCAGTG ATAGTTACAA TAGTGTGTT GAATGTTGAA CGGATGTGTG CTACACCATT 3960  
 TTCAATATTC TTTTCACTC TACGTTTACG AGATACTTGT TTACGTGCCA TTTAAATTT 4020

55

EP 0 786 519 A2

	CGCGCGTkgT	TTTTCGTTTT	TTGACCACGA	ACTGGTAAAC	CACGACGGTG	ACGGATACCC	4140
	ACCGTATGAT	GAAATTnCCA	TTAAAcGTTT	GATATTTAAG	TTAGTTTCAC	GACGTAAGTC	4200
5	ACCTTCGACT	TTATAACCGT	CTACAACCTC	ACGGATGCGA	CCTAATTCGT	CATCAGTTAA	4260
	ATCTTTCACA	CGAGTATCAG	CTGATACGTT	AGCTTCTTCA	AGAATTTTTT	GAGCAGTTGA	4320
	CGTACCGATA	CCGTATATAT	AAGTTAATGA	GATAACTACG	CGTTTTTCAC	GTGGAATATC	4380
10	TACTCCTGCA	ATACGTGCCA	TATTAATTTA	CACCTCTCTT	TTATTAACCT	TGTCTTTGTT	4440
	TGTGTTTTGG	ATTTTCACAA	ATTACCATTA	CTTTACCTTT	ACGTTTAATG	ACTTTACATT	4500
15	TTTCGCAAAT	AGGTTTTACT	GATGGTCTTA	CTTTCATTTT	TATACCTCCC	TATATTATGG	4560
	AGTGACGATT	ATTTATAACG	ATAAGTAATT	CTTCCGCGTG	TTAAATCGTA	CGGAGACATC	4620
	TCAACAGTTA	CTTTGTGCGC	AGGTAGAATA	CGAATGTAAT	TCATTCTGAT	TTTACCACCT	4680
20	ACGTGAGCnA	AAATCTCATG	ACCATTTTCT	AATTCTACTT	TAAACATTGC	GTTCCGTAAA	4740
	GTATCTAATA	CAGTACCTTC	TAATTCAATT	ACATCTTGTT	TAGCCATTGA	TTAACTTCCC	4800
	CCTTTTTGCA	ATAGTAAGGT	AATCGTCAAT	AGACAACCTT	ATTGTTACGA	ATCTATCAGT	4860
25	GATTAATTTT	ATAAGTTAAA	CAAAAATTAC	GGGAATTAAT	TATCGTTAAT	TGCCACTCTC	4920
	ATCTATCTAA	TATGATTAAA	TCATGCCTCA	CTTAAAATAG	ACCGCTAAAA	GTTGATCTAT	4980
	TACAAATGAT	CTAAAATATC	AATGACATCT	TTGGTAACGT	CGCTAATATC	TTTTGAACCA	5040
30	TCAATATTTT	TCAATACACC	TTTTTGATCA	TAGAAATCTA	AAATAGGCTT	AGATTGTTTA	5100
	ATATTAACAC	TCAAACGATT	AGCTACCGTT	TCAGGATTAT	CATCTTCTCG	TTGATACAAT	5160
35	TTACCACCAT	CGATATCACA	AATACCTTCG	ACTTCGGAGG	ATTAAATACA	AGATGATACG	5220
	TTGTACCACA	TGACTCACAG	ATTGACGAC	CTGTAAGACG	GTTCATTAAAT	TCTTCTTCCG	5280
	GAACTTCGAT	ATTGATGACA	GCATCAATGT	TTCTGTCAAG	CTCAGACATA	ATATTATTTA	5340
40	ATGCCTCAGC	TTGCTCGATT	GTTCTTGGA	AGCCATCTAA	TAAAAAGCCT	TTTTTTGCAT	5400
	CGTCTTCAGA	AATTCTTTCC	TTAACGATAC	CTACAGTCAC	TTCATCAGGA	ACTAATTTCG	5460
	CACGGTCCAT	ATAAGACTTA	GCTTCTTTAC	CTAATTCAGT	TTCTTCTTTT	ATAGCTTTTC	5520
45	TGAACATGTC	ACCAGTTGAA	ATGTGGGGTA	TTGGGAATTT	CTTGaCAATT	TCACTTGCTT	5580
	GAGTTCCCTT	ACCTGCGCCA	GGTAAACCCA	TCAAATGAT	ATTCATAAGT	GCCCTCCTAA	5640
	AATTATCTAC	CACCAAAGCC	TTTATATTCT	TTTTGAGATA	CTTGAGCTTC	TAAAGATTTT	5700
50	ATTGTTTCAA	TCGCTACACC	AATAACGATA	AGTAACTTG	TACCACCAAT	CTGAATTGAT	5760
	TGTGGTAATC	CCATAAACTT	AGTTGCTAAT	ATCGGTAGAA	TTGAAATAAC	GGCTAAGAAG	5820
55							

	CCAGGTCTAA TACCTGGAAC ATAGCTACCT TGTTCCTTAA GGTATCAGC CATTTTTTCC	5940
	GGATTAACCT GTACAAATGC ATAGAAGTAT GTGAATAGTA TAATTAGTAC AATATATACA	6000
5	ACCATACCAA CATTACTTGA AGGATTGCA GCATTGCAA TGTTCGTGC CCATTCTTTA	6060
	TCTGGATAGA ACAACGTAA TGTCTAGGC AGTAAGAAGA ACGCCATTGC AAAGATTACA	6120
	GGAATAACAC CGGCTGAGTT CACTTTTAAA GGAGATAAG TTGCCTGTGA ACCTAATCTT	6180
10	TGAGCAGTTT GTTCTTAGC ATATTGAATC GGAATTTTAC GAACGGCTTC AAGTACATAA	6240
	ATAGCACCTA CAGTTAATAG TATCAGTGAC ACTAAAAGTC CTAATACTTT CAACCATGCT	6300
15	AATGATGTAT CTTCTTGCCC AACGAACGCA TTTGTCCAAA TTGAATTAGA CTGGCTGGCA	6360
	ACGTTGATAA AATACCGCA AATATGATAA TAGAAATACC ATTACCAACA CCGAACTGAG	6420
	TGATTTGATC ACCAAGCCAT ATTAAGAAAG CAGTTCCTGC TGTnCAAAAC TAGTGCTATT	6480
20	AATAAATAAC TCATAATTGA CTGATTGATA ATCAGCGCAC CTTTGAGATA ATTATTAAAT	6540
	TGGAATGCCA TACCTATAGA TTGGATAAAT GCTAAAGAAA TTGCTAAATA ACGAGTAACG	6600
	TTATTTAACT TTCTTCTACC TACTTCACCT TGTTCGCCC ATTCTGAGAA TTTAGGGACA	6660
25	ATATCCATT GTATAATTG CATTACGATT GATGCAGTGA TGAGGGTAC AATACCCATT	6720
	GCAAAAATAG AAAATCGTTT CAAGGCTCCG CCACCAAAAG TATTTAATAA CTCAGTGGCA	6780
	CCTTGAGAAC CTTGGGGATT ATCAAAAGCT GCAGGATTTA CTCCTGGAGC TGGTATATAA	6840
30	GTCCCTATTT TAAAAATTAC TAACATTGCT AGTGTAAGA AAATCTTGTT ACGAACCTCT	6900
	TTTGTTCTAA AGAAGTTCAC AAGGGTTTGA ATCATTAGAT CACCTCGTGT GCTCCACCTT	6960
35	TAGCATCAAT AGCTTCTGCT GCTGAAGCTG AGAATTTATG AGCTTTCCT GTCAATTTCT	7020
	TATCAAGTGA ACCATTACCT AGTATTTTGA TACCAGATTT TTCATTCTTA ACAACACCAG	7080
	ATTCTACTAA TAAAGCTGGA GTTACTTCAG TACCATCTTC AAATTTATTA AGTTGGTCTA	7140
40	AGTTAACAAT AGCATATTCT TTACGATTTA TGTTAGTAAA ACCACGTTTT GGTAAACGAC	7200
	GGAATAATGG TAATTGACCA CTTTCAAATC CTGTCCTTAC ACCACCGCCT GAACGAGCTT	7260
	TTTGACCTTT GTGTCCGCGA CCACTTGTTT TACCGTTACC TGTCGCAACA CCACGTCCAA	7320
45	CACGATTGCG TTCTTTACGT GAACCTTCTG CCGGTTTTAA CTCATGTAAT TTCATTTCGG	7380
	CACCTCCTTG ATTATTTTTC TTCTACTGTT ACTAAGTGCT TAACTTTGTT GATTGCCCCA	7440
50	CGAATAGCAG GGTTATCTTC AACAACTACT GAACTGTTAG TCTTTTAAAG ACCTAAAGCT	7500
	TCAACAGTTT TACGTTGTGT TTCAGGACGA CCAATAACAC TACGAGTGAG GGTAATTTGT	7560
55	AATTTAGCCA TAACTAGTTT TCCCTCCTTA ATTGTATAAT TCTTCTACTG TTTTGCCACG	7620

EP 0 786 519 A2

	CATGTTGATT GGTGTGTTTG ATCCTAATGA TTTACTTAAG ATATCAGTGA TACCTGCTAA	7740
	TTCAAGTACG GCACGAACAG GACCACCAGC GATAACTCCT GTACCAGGTG CAGCCGGTTT	7800
5	CATAAATACG CTTCTGAAC CGTAACGGCC AGTAATTGTG TGTGGAGTTG TACCTTCAAC	7860
	ACGTGGAACA ACTACTAAAT CTTTTTTAGC TGCTTCAACA GCTTTTTTGA TTGCTTCTGG	7920
	TACCTCTTGA GCTTTACCAG TACCGAAACC TACACGACCA TTTTGTCTC CAACTACAAC	7980
10	TAATGCAGTG AAACGGAAAC GACGACCACC TTTTACAAC TTTGCTACAC GGTGATTGT	8040
	AACAACGCGT TCTTCAAAT CTTTCGTCTC TTCTTCTCTA CGAGCCATGT ATTTGTCCCT	8100
	CCTTTAAATT AAAATTCTAA TCCGCTTTCT CTTGCTGCTT CAGCTAATGC TTTAACACGT	8160
15	CCGTGATATA AATATCCTCC ACGGTCAAAT ACGATTTCTT TAATGCCTTT GTCAGCAGCT	8220
	TTTTTAGCAA TTGCTTCACC GACTTTAGTT GCTAATTCAA CTTTAGTTGC TGTAGTAGCA	8280
20	ATGTCGCTGT CTTTTGAAGA AGCTTGAGCT AATGTTACGC CTTTATTATC ATCAATAATT	8340
	TGAGCGTAGA TATGCTTGTT TGAACGATAT ACGTTTAAAC GTGGCTTTTC AGCTGTACCT	8400
	GATAAGTTAG TACGAACACG AGCATGTCTT TTTAAACGCA CTTTATTTTT ATCAATTTTA	8460
25	CTGATCATTT CAATACTCCT TTCTTTAGAG TTTATCTATT ATTTACCAGT TTTACCTCT	8520
	TTACGGCGAA CGTATTCACC TTGGTAACGA ATACCTTTAC CTTTGTAAGG CTCTGGAGGT	8580
	CTTACTGAAC GGATGTTAGA TGCTAATGCT CCAACTTGTT CTTTTGAAAT ACCTTCAACT	8640
30	TTAACGACTG TGTTTTTCTC AACTGAGAAA GTAATGTTTT CTTCAGCTTT AATTTCTACT	8700
	GGGTGAGAAT AACCAACGTT AAGGATTAAG TCTTTACCTT GCATTTGAGC ACGGTAACCT	8760
	ACACCAACAA GTTCAAGTAC TTTTACGTAT CCTTGAGAAA CACCTGTAC CATATTGTTT	8820
35	AATAAAGCAC GAGTTGTACC ATGGTTTGTT CTATCTTCTT TAGAATCAGA TGGTCTTACA	8880
	ACTTCAATTG TGTTTTCTC TTGTTTGAAT GTCATTCTTT CATTTAAAGT TCTTGATAAT	8940
40	TCACCTTTAG GACCTTTAAC AGTTACATGA TTTCCATCAA AAGTTACTGT TACGTCACTA	9000
	GGGATGTCAA TAATTTTCTT ACCAACACGA CTCATGTTAT GGCACCTCCT TATTTTTTAT	9060
	TACCAAACGT ATGCGATAAT TTCTCCACCA ACATTACGTT TTCTTGCTTC TTTGTCAGTG	9120
45	ATTACACCTT CAGAAGTTGA TACTAATGCA ATACCTAAAC CATTTAATAC TTTAGGCATT	9180
	TGCTAGCTT TTGCATAAAC ACGTAAACCT GGTTTTGAAA TACGTTTTAA TCCTGTGATA	9240
	ACACGCTCAT CGTTTTGACC ATATTTTAAG AATAAACGAA GTACACCTTG TTTATCATCT	9300
50	TCTACGTATT CAACATTTTT AATGAAACCT TCACTCTTTA AGATTTCAGC AATTTCTTTT	9360
	TTAATATTTG ATGCAGGTAA TTCTAACTTC TCGTGACGCA CCATGTTTGC GTTTCTTACA	9420

55

	TCTTTTTTAT TACCAGCTAG CTTTACGAAC GCCAGGGATT TGGCCTTTGT AAGCTAATTC	9540
	ACGGAAACAA ATACGGCATA ATTTAAATTT ACGATATACA GAATGTGGAC GGCCACAACG	9600
5	TTCACAACGA GTGTATTAC GAACTGCATA TTTTGTGTTT TTTTGTGCT TAGCAACCAT	9660
	TGAAGTTTGA GCCACTTAAT TAGCCTCCTT TAAATAATTA TTTACGGAAT GGCATACCGA	9720
	AGTTAGCTAA CAATTCACGA GCTTCTTCAT CAGTGTTAGC AGTCGTTACG ATAACAATAT	9780
10	CCATTCTCT AACTTTACTT ACTTTATCAT AGTCGATTTC TGGGAAAATT AATTGTTCTT	9840
	TAACACCTAA AGTGTAGTTA CCGCGTCCGT CAAATGCTTT TTTAGAAACA CCTTGAAGT	9900
15	CACGTACACG TGGTATGAT ACTGAAATTA ATTTGTCTAA GAATTCATAC ATTCTTTCAC	9960
	CGCGAAGTGT TACTTTCGCA CCGATTGGCA TACCTTCACG TAAACGGAAA GTCGCGATTG	10020
	aTTTTTTAGC TTTAGTTACT AATGGtTTTT GACCAGTGAT CAATTCTAAT TCTTCAACAG	10080
20	CATTGTCTAA TACTTTAGAA TTTTGTACTG CGTCACCTAC ACCCATGTTT ACAACGATTT	10140
	TATCTATTTT TGGTACTTCC ATTACTGAAC TATAATTGAA TTTTTCATT AAGTTTTCAG	10200
	TAACCTCAGT GTTaAACTTT TctTTTaAAC GGTTCaAAGT GGGATCCTCC TTTCaACTTG	10260
25	TtATTAATTA TTAGAkTTAA TTTCTTCGCC AGATTTTTTA GCGATACGAA CTTTTTTACC	10320
	ATCAACAAAT TTGTAACCTA CACGAGTTGG TTCGTTTGTT TTAGGGTCCA ATAATTGTAC	10380
	ATTAGAAACA TGGATTGCTG CCTCTGTTTC TAAGATTCCA CCTTCAGGAT TTAATTGAGT	10440
30	TGGTTTTTGG TGTTTTTTCA TAATGTTAAC ACCTTCCACA ACGACACGGT CTTTTTTAGG	10500
	TAGAGTAGCA ATTACTTTAC CTTCCTTACC TTTGTCTTTA CCTGCGATAA CTTTAACGTT	10560
35	GTCACCTTTT TTGATATGCA TGTGGGCACC TCCTTATTTG TATTGGTTGT TATTAAATTA	10620
	AGTACTTCTG GTGCTAATGA TACGATTTTC ATGAAGTTAC CTTACGTA TACACGAGCA	10680
	ACAGSTCCGA AGATACGAGT ACCACGTGGG CCTTTGTCTAT CACGGATGAT AACACATGCA	10740
40	TTTTCATCAA ATTTGATGTA TGAACCGTCA TTACGACGAA CACCTGACTT AGTACGTACG	10800
	ATTACAGCTT TGACAACGTC ACCTTTTTTA ACAACGCCAC CTGGTGTGTC ATTTTTAACA	10860
	GTACATACGA TAACATCGCC GATGTTTGCT GTTTTACGAC CAGATCCACC TAATACTTTG	10920
45	ATTGTAAGAA CTTACGAGC ACCAGAGTTG TCTGCTACTT TCAAGCGTGT TTCTTGTGG	10980
	ATCATTAGTT AAACCTCCCT TATCTCTAAA CTTGTATTAA ATAATTACTG ACTCTTCAAC	11040
	AATCTCTACT AAACGAAAAC GTTTTGTGTC TGATAAAGGA CGAGTTTCIT GAATTTTAAC	11100
50	AATGTCTCCT AATTAGCTG AATTGTTTTT ATCATGAGTT TTGTATTTT TAGAGTATTT	11160
	TACTCGTTTA CCGTATAATT TGTGTGTTTT GTAAGTTTCA ACAAGTACTG TAATAGTCTT	11220
55		

	TTTTGTAACC TCCTCTTACT TAATTATTGA TTAGCCTTAC TTGTTCAAT TTCTCTTTCA	11340
	CGAGCAACAG TTTTLAGACG TGCAATCGTT TTTCTTACTG TACGAATACG TGCAGTTTCT	11400
5	TCTAATTGAC CTGTAGCTAA CTGAAAGCGT AGGTTAAAAA GCTCTTCTTT TGAAGATTG	11460
	ATTTGTTCTT CGATTCTGA AGTGGTTAAG TCTCTAATTT CCTTAGCTTT CATTTGTTTC	11520
	ACCACCCAAT TCCTCACGTT TTACAAACTT AGTTTTTACT GGAAGTTTGT GACTTGCTAA	11580
10	ACGTAGTGcT TCACGCGCAA CTTCTTCAGA AACGCCAGCA ACTTCGAATA AAATTCTACC	11640
	TGGTTTAAACA ACTGCGATCC AGCCTTCAAC CGCACCTTTA CCAGCACCCA TACGTACTTC	11700
15	TAAAGGTTTT TTAGTATATG GTGTATGTGG GAAGATTTTA ATCCAAACTT TCCCGCCACG	11760
	TTTCATGTAA CGTGTcATTG CTATACGAGC AGATTcGATT TGACGAGATG TGATCCAAGA	11820
	CGTTGTTGTA GCTTGTA AAC CAAACTCACC AAATGTTACG TAcTACCGCC TTTAGAACGA	11880
20	CCAGTTGTTT TAGGACGATG TTGACGACGA TATTTTACAC GTTTTGGTAG TAACATTATT	11940
	ATTTTCCTCC TCCACTAGTG TTCTTAGTAG GAAGAACTTC TCCACGATAA ATCCATACTT	12000
	TAACGCCTAA TTTACCGTAA GTAGTGTcAG CTTcAGCGTG cGCATAATCG ATGTcAGCAC	12060
25	GTAACGTATG AAGTGAACA GTTCCTTCTG AATATTGTTc AGCACGAGCG ATGTcAGCTC	12120
	CGCCTAAACG ACCAGATACT TGaGTTTTGA TACCTTTAGC ACCAAGTTTC ATAGCTCTAG	12180
	TGATTGCTTG TTTTTGTACA CGACGGAATG AAGCACGGTT TTCTAATTGA CGTGCGATGT	12240
30	TTTCAGCTAC TAAACGAGCG TCAAGATCAA CTTTTTTGAT TTCAATTACG TTGATGTGTA	12300
	CTTTTTTATC AGTTAACGCA TTTAATTGTG TGCgTAATTT TTCGATTTCT GAACCGCCTT	12360
	TACCAATTAC CATACCAGGT TTACCAGTAT GAATTGCAAT GTTGATACGG TTTGCAGCAC	12420
35	GTTCAATCTC TACGTGAGAA ACTGATGCTT CTTTTAATTC ATTATCAATA AATTTACGGA	12480
	TTTTTAAATC TTCGTGTAAA AGTGAAGCGA AGTCTTTTTc AGCATACCAT TTAGCTTCCC	12540
40	AATCACGGAT AATACCAACA CGAAGTCCGA TTGGATTAA TTTTcGACCC ACAGTATTCC	12600
	CTCCTTAAAA GTTAATTAAG CTTCTTTAGC TTCTTCTTTA CCGTCACTTA CGACGATTGT	12660
	AATGTGGCTT GTACGTTTTGT TAATCGCACT TGCACGACCT TGCGCACGTG GACGGAAACG	12720
45	TTTTAATGTT GGTcCTTCGT TAGCATATGC TTCTTTAACT ACTAATTCAT CTGTGTTcAT	12780
	GTCATAGTTA TGTTcAGCAT TAGCTAAAGC GGACATTAA TCTTTTTCAA TTACTGGTGA	12840
	TGAAGCTTTG TTTGTTAATT TTA AAATTGC AATAGCTTCA GCAGCATTTT TACCTCTGAT	12900
50	TAAGTCAAGA ACTAGTCTTA CTTTACGAGG TGCGATTCTT ATTGTCTAG CAACCGCTTT	12960
	TGCTTCCATT AGGATGTcCT CCTCTACTTA ATAGATATTA TCTTCTTGTT TTCTTGTCGT	13020

55



	TATCTTCAGT TACATATACA GGTACGTGTT TACGTCCGTC GTATACTGCA AAAGTATGTC	13140
	CGATGAAATT AGGGAAAATT GTAGAACGAC GTGACCATGT TTTGATTACT TGTTCCTTTT	13200
5	CGCTTCCTTC TTGAGCTTCA ACTTTTTTCA TTAAATGCTC ATCGACGAAA GGTCCTTTTT	13260
	TAATACTACG AGCCATTTGG GCGCCTCCCT TCTTATTATG TCGTGCAGC TTTAAGCCGC	13320
	ACACCCAAAT AAGTTGATTA TATTATTTTT TCTTACGTCC ACGAACGATA AGTTTGTCTG	13380
10	ATGATTTTTT ACCACGACGA GTTTTCTTAC CAAGCGTAGG TTTACCCCAT GGTGACATTG	13440
	GAGATGGTCT ACCGATAGGA GCACGACCTT CACCACCACC GTGTGGGTGA TCGTTAGGGT	13500
15	TCATTACAGA ACCACGAACT GTTGACGGA TACCTTTCCA TCTTGAACGT CCGGCTTTAC	13560
	CAACGTTAAC TAATTCGTGT TGTAGGTTAC CAACTTGACC GATTGTAGCA CGGCAAGTAG	13620
	ATAAGATCAT ACGAACTTCA CCAGATCTTA ATCTGATTAA TACGTATTTA CCTTCTTTAC	13680
20	CAAGTACTTG AGCACTTGCA CCAGCTGAAC GAGCGATTG TCCACCTTTA CCAGGTTTAA	13740
	GCTCGATGTT GTGTACTACT GTACCAACTG GAATGTTTTG TAATGGTAAT GCGTTACCAA	13800
	CTTTGATGTC AGCTTCAGCA CCACTTTCAA CGATTTGACC TACTTCTAAT CCTTTAGGAG	13860
25	CAATGATATA TCGTTTTTCA CCGTCTGCAT ATACAACTAA AGCGATGTTT GCTGAGCGGT	13920
	TTGGATCATA TTGAATAGAA TCAACTTTTG CATTGATACC ATCTTTGTTA CGTTTGAAAT	13980
	CGATAACACG GTATTGACGT TTGTGTCCAC CACCATGGTG TCTTACAGTC AATTTACCTT	14040
30	GGTTGTTACG TCCCGCTTTT TTCGGTAGCG GTTTTAATAA TGACTTTTCA GGTGTAGTTT	14100
	TCGTGATTTT TCGGAAATCT AACGAAGTCA TATTACGACG ACCATTGTGT ATTGGCTTAT	14160
35	ACTTTTTAAT AGCCATTGTC GCTTACCTCC TTAATGGTAA TTGTTTTATT AGTTAAATAA	14220
	GTCGATTGAT CCTTCTTTAA GAGTTACAAT CGCTTTTCTT CTTTGTGTTG TATAGCCTTG	14280
	GTAACGGCCC ATACGTTTTT TCTTAGGTTT GTAATTCATG ATATTAAACAC TTGCAACTTT	14340
40	TACGTTGAAG ATTTCTTCAA CTGCCATTTT TACTTGTGTT TTGTTAACAC GAGTATCAAC	14400
	GTCGAAAGTG TATTGTCTT CAGCCATTGC TTCAGAAGAT TTCTCAGTGA TTACGGGGCG	14460
	CTTAAGAATA TCTCTTGCTT CCATTATCCG AGCACCTCCT CAACTTTTTT AGCAGCAGCT	14520
45	TCAGTAATTA CTAAGCTGTC AGCATTAGTG ATATCTAAAA CATTTAAACC TTGAGCAGTT	14580
	GTCACTTGAA CGCCAGGGAT GTTGCGTGCT GATAATTCAA CATTTACATC TTCGTTTTCA	14640
	GTAAC TACTA ATACTTTTTT AGGTTGTTCT AATGTAGATA ATACATTTTT GAATTCCTTA	14700
50	GTTTTTGGAG CTTCAAGTT GAATGCGTCA ACTACAGTTA AGCCATTCTC TTGAGCTTTG	14760
	AAAGATAATG CTGAGCGTAA AGCTAAACGA CGCATTTTCT TAGGCATTTT GTATGCATAA	14820
55		

	CCTTGACGAG CACGACCTGT TCCTTTTTGC TTCCATGGTT TACGTCCGCC ACCGCTTACT	14940
	GCTGAACGAT TCTTAACAGC ATGCGTACCT TGACGTAATG AAGCACGTTG TAAATTAATA	15000
5	GCTTCGAATA AAACGCTATT ATTTGGCTCA ATACCGAATA CTGCATCGCT TAATTCGATT	15060
	GAACCTGATT TAGTCCGTC TAATTTTAAA ACATCATAAT TAGCCATTAT GCATTTCTC	15120
	CTTTCACCTC TTATTATTTA TTACCTTTTT TAATTGAAGT TCTGATTTCT ACTAAACCTT	15180
10	TTTTAGGTCC AGGTACGTTA CCTTTTACTA AGATAACTTT GTTTTCTGTG TCAACTTGAA	15240
	CTACTTCTAA GTTTTGAACA GTTACAGTGT TTCCACCCAT ACGTCCTGGC ATTTTTTGGC	15300
15	CTTTAAATAC TCTAGAAGCA TCTGAAGCCA TACCTACAGA ACCTGGTGCT CTGTGGAAAT	15360
	GAGAACCGTG TGACATAGGT CCACGAGATT GTCCGTGGCG TTTAATTGCA CCTTGGAAAC	15420
	CTTTACCTTT TGATACGCCT GTTACGTCAA TAACGTCGCC AGCTACAAA GTATCTACTG	15480
20	AGACTTCTTG AaCCTAcTcG TAAGCATCCA CGTCTACATT GCGGAATTCA CGAATGAAGC	15540
	GCTTAGGTGC TGCGTCAGCT TTTTAGCGT GACCTTCAGC TGGTTTATTA GCATATTTAT	15600
	TAGATTTTGC ATCTTTTTTG TATGCTTTTT TGTCTTCAA TCCAACCTGG ATTGCGTTGT	15660
25	ATCCATCAAC TTCTACAGTT TTCTTTTGTA ATACAACATT TTCTTTAGCT TCTACTACTG	15720
	TTACAGGGAT TAATTCACCG TTTCTCCGA ATACTTGTGT CATCCCAATT TTTCTTCTA	15780
	AGATTCCTTT GGTCAATCGAA AGTCCACCTC CTAAAATTGT CTATTATAAT TTGATTTCGA	15840
30	TGTCTACACC AGATGGTAAG TTTAAGCCCA TTAAAGCGTC AACTGTTTTT GGTGTTGGGT	15900
	TTACAATATC GATTAAACGT TTGTGTGTAC GTTGTTTCGAA TTGTTACAGT GAATCTTTAT	15960
35	ACTTATGCAC GGCACGGATG ATTGTGTAAA CTGATTTCTC AGTTGGTAAC GGAATTGGTC	16020
	CAGAAACATC TGCACCAGAA CGTTTCGCTG TTTCTACAAT CTTCTCTGCT GATTGATCAA	16080
	TTACGCGGTG ATCATAAGCT TTTAATCTGA TTCTGATTTT TTGTTTTGCC ATAATTTTCC	16140
40	CTCCTTATTC GTCTACATTT AGTGATAGAC TTCTCCACGA AAATATCTT ACACAGCGCC	16200
	ATGGCAAAGC GGCCGGGTGT GTCAGTAACC TTTCGCTTCA TCGCTTTTCT TAAAGTCCAA	16260
	CGTTAGTTAT ATTACACGAA AAACATCGAT AAATCAAGC TTTTCACATA ATTTTCTAT	16320
45	CTGTCTAACA CATACTTTTA TATTTnACTT TATATACTTA GTCAGTTCAA CTATTTTCGA	16380
	GATATTTTnA ATTTCCn	16397

(2) INFORMATION FOR SEQ ID NO: 206:

50

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29555 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

5	TnAGTTGTTT CTGCCACGAA AGATTCAATG GCTTTTCTTG CTTTACGCTT TTCTTTCAAT	60
	GGCAAATCAC CAATCATTTT TTTAAGGTGA TGTGGGTTTA CAACACCACT ATACTGGTAG	120
	TCATTGAAAn TTGTTTTTAG GGCTTGTTCA TCGATAGATC TCTCTCCAGC AAATCCTTTG	180
10	AACTCCGCTT CTTTTTTAAT ACTTTCGAAA TTAACATATT CTTGATCGAT ATCATCATCC	240
	TTATTTAAAG AAGGTACAAC ATTGTCGATG AATTCTCTAA TTAGATCTCG TTTTAACCTC	300
15	AATGtCGGAT CATCTGCATG ATCTAAAATG CGTCTAATTT GTTCTTGgTT ACGACGTTGT	360
	TCCGCTTTGT CTTCAAGATC AATTGTCTC AATATATTCA TAATATAATT CACATTAATC	420
	GTATCATTAC GCATCATTTTc TATTTCGAAA TCAATATCAT TTAAAATGGA TACTTTATTT	480
20	TTCTCAGCCG TCGCTCTTTT TACTTGATCG TACACAGCTA AATATTTACT TTTATAGTCT	540
	TCATTCTCTT GTTCATCCAT TCCAATTTCA TCAATTGTAA ACTCAAACCTC GTCAAATGCT	600
	TTTAAACGTA ATATTATTTT AGCTAATAAA CGATAAGCtT CAACAAAGCG CTTTAGCTCT	660
25	TCTTCATCtT GaATGtCATC AACCATGTGT GGTGTGGCA CAATCATTTT AAGCTCACGA	720
	TAAGCGTCCA TAAATTCCTT TTTATACTCT TCATAACTGC GCATTAAAAT TGTATCCGTA	780
	TCATTGTTTT GTGAGAATAC TCTCAGTGCA TCGTCTGTCT CTTTTTTCAA GTCACGATAG	840
30	TTTACAATTT TACCAAATGG CTTTGATTCT TTTTCAACCC TATTGTACG TGAATACGCT	900
	TGAATTAAAT CATGATACAT TAAATTCCTA TCAACATATA AAGTGTTcAG TACTTTACTA	960
35	TCAAAACCAG TTAAGAACAT ATTAACAACG ATTAAGATAT CAATTTTACT ATCTTTAACG	1020
	CCCTTTTTAA CGTTTTTtGA AATATGATTA AAATACTCAT TAGTTGtGGC TgNtGaAAAA	1080
	TTCGTCTCGA ACTTTTTATT ATAATCACTA ATCATTATCT CTAATTTTTc ACGTGAATGA	1140
40	TATGGCACTT CACCATCACG ATCATCTTCA TTAGGTTTAA ACGTAAATAT ACCAGCTATC	1200
	GTTAACGGTT GTTCCAACCTT TTTGTTAAGT CGCTTAAATG TCTCATAATA TTTAATAAGC	1260
	GCGTGAATAC TTTGGACTGT AAATATACTT GAATATTGAC GATTACGTGT ATATTTATCA	1320
45	TGATTATTGA TGATATGTCG TGTTACTAAT TCCACACGTT TATCCGCTAA CCATACTTCT	1380
	TCCGTATCAA TTGCTTCAAC CATGCTGTtA TCTTCTGCTT TTAAAGCTTT ATTTTTAAAA	1440
	GTATTAATAT AGTCAACTGA GAAACCAAGT ACATTACCAT CATGAATGGC ATCTCTAATT	1500
50	AAATACGTAT GTAAGCATCT ACCGAAAATA TCTGCAGTTG TTCTACCATC TTGACTACTA	1560
	TTTTCTGGAA AACGTGGCGT ACCAGTGAAT CCAAAGTATT GGGCATTTTT GAAATGTTGT	1620

55

	ACTTTATTTCG TTTTATACTG TTCTAATAAA GGGGCATTCC CTTGAATCGC TTTAGCCATT	1740
	TTTTGAATCG TCGTTACAAT AAGTGGCAAA CTTTTATCAT TTAGTTGGCG TACCAGTTGC	1800
5	GAGGTATTAA AAGTTTTGTC TACAGCACCC TTAGCAAATT TATTAAATTC CTCTTCTGTT	1860
	TGACTATCCA AGTCTTTACG GTCAACCAAA AAGATAACTT TCTTAATGTC ATCTTGCTGT	1920
	GATAAAATCT GACTCGCTTT AAAAGAAGTC AACGTCTTAC CACTTCCAGT TGTATGCCAT	1980
10	ACATATCCAT TATTCCCTGT CTCAGTCGCT TGTGAATAA GTGCTTCTAC CGCATACACT	2040
	TGATACGGAC GCATTGCCAT CAGTATTCTA TCTGTTTCAT TAATAATCAT ATAGCGCGAT	2100
15	ATCATCTTAG CTAATTGACA AGGTCTCATA AATGACTCAG CAAACGATTG CAATGTATTG	2160
	ATACGGTTAT TCTGTTTATC ACTCCAATAA AACATGTGAC TCTTCAATAG TTCGCTATCA	2220
	TTATTAGAAA AGTATCGCGT TTCAACACCA TTAATAATGA TAAACATTG TATGTAGCGG	2280
20	AATAAGCCTG TGTAATTTTG TTTGCGGTAA CGTTTTACTT GGTAAACGC CTCATTAATA	2340
	TCAATACCTC GACGTTTCAA TTCAACTTGG ACAAGGGGTA GTCCGTTGAT TAATATCGTT	2400
	ACATCATAAC GTGCTTTATA TGTATCCTCG ACAGATACTT GATTCGTCAC TTGAAACTTA	2460
25	TTTTTACACC AACTTTTCGT ATCTAAAAAC GACAAATAAA TCTCAGACTC ATCATCACGT	2520
	CTAAGTGGTA ATTTATCACG TAAAATACGG GCACTCTCGA AAATACTTTT TCCATCAATC	2580
	ATCGTTAACA GACGTTGAAA TTCTTTATCT GTTAAGGGAT TGCTTCTAA TTTGTCCGCA	2640
30	TGACGCTCAT TTAAATTCGT TCTAAAATTA TCAAGCAATT GCTTATTATC ACGTATCGTT	2700
	ACTCTTTCGT AACCCAATTG TTCAAGTTGA TTCATCATTT CATTTTCTAA TGCGTATTCA	2760
	CTTTGGTATG CCATTCATAT CCCCTTCCAT ACACTTTCTA TTGCTCTAAA TATATCATAA	2820
35	ACTTTAATGA AAAATGTTTG TTTTTATCT TCAAACGTAA ATTTATTCTA ATTTTATTGT	2880
	CTTATCTTTT AATATTTGTC TTTGAGGTAA GTCGTATACT AAAATTTGAA TACAAATAAT	2940
40	CAAAATCATTG ATAAATTTTT TGTCTACGAT TAATGGAGGG ACTTGAATGG TGTTAATTAC	3000
	CTATCAAATC ATTTTATTTT TTATTATTAG TCTAAGTTAC TATTTAACTT TAAATCATT	3060
	CATGGCAGTC ACTGTAGGTA ACTTCAC TTC AATATTCGGC ATGTTTCGCAG CCATACTCTT	3120
45	TATGTACTAC TACCTACTCT ATAAAAGTCC CGAATACAAT CAACGCAAAC GATTTAAACA	3180
	TTTCATTCAT ATCACTAATT TGATAATAAT TGCTTTTAGC ACCTTCGTAT TAGTTCATTT	3240
	AGCATTAAAA TTATTCTTCA GCATTTAATT TCCATCTATG AAAAAAGCAA AGCTCAAATC	3300
50	TGAACTTTGC TTAAATTTGT CACGCCTTTA TCATTTTCAA AATAGCCTCT ATGCCAGTTT	3360
	TACAACTTG TAGCAACAAT TTTTCATCAA GCAACTGAAT CACATCAAAA ACTTCAATTG	3420
55		

	GTCGCAAGAT	GCTTCCTGTA	ATTATCAAGT	GCCATTTTCG	ATTGGGTTAT	ACAATCTAGA	3540
	ATCGCATGAT	AATTTAATGC	TACAAATCGA	TAGTACAATA	TATCTACCGT	GAATAACTGT	3600
5	GCAAATAGTG	ACGTTGTAGC	CGCCATACGC	ATTTTCATTTT	CATCAGTTCT	GCCATAAATC	3660
	AATGCATAGT	CTGCAATTTG	AGCCACTGGA	TTATTAGCTG	TACTAGATAT	AGTTATGATG	3720
	GGAATACTGT	AATGTGTGGC	CACCTGTGCA	ATTGACTGCA	ATTCACTATG	ACTACCTTGA	3780
10	TTCGTCACAA	AAATCATGCA	ATCTCTATCA	TCATGCGTCG	CAAATGTTGA	CACAAGTAAA	3840
	TGCGTTTCAT	GTAATAACCT	GACATTTAAG	CCAATACGAG	ATAACTTTTG	AAAAAGATCA	3900
15	CCAATAGTCA	AACTCGATGC	GCCAAATCCA	AATAAAAATA	TTGTCCTGGC	ATTTTTCAAC	3960
	ACATCACAAA	TTGCATCAAT	TTGCGCATCC	ATAATATTAG	TAGCTACAAA	TCGCATCGTA	4020
	TTCGTTGCTC	TAGCAATCAT	TTTATTTTTC	AAAGTTTCTA	CAGATTTCATT	TTCAATCAAT	4080
20	TCTAAATGTG	GATTGGTTGC	AATATCTTCG	GGTAAGTATC	GAGATATCGC	AATCTTTAGC	4140
	TCTTGAAAAC	CTTGATGTGT	CATTTTCCGA	CTAAATCTAA	CAATTGATGC	TGTACTAACA	4200
	TTCGTAACAT	CTGCCAAATC	ATTCACAGTC	ATATCAATGA	TTTTATGTGG	ATTCTTTAAA	4260
25	ATGTAATCAG	CGATTATCTT	TTCTGTCTTC	GTAAAATCAC	TCAACTGCTT	ATCAATGCGA	4320
	TATAAAATAT	TTGTCATCAT	TAATCACCCA	ACAAATCTGT	CTGTCGCATC	GCCTTTGTGC	4380
	TTCCAAATAA	ATATGTACAA	ACGAATCCAC	CAGCATACGC	AGCAAGTAAT	CCTGCAATAT	4440
30	AACCTAAATA	CATATTATCT	GAGATTAATG	GTAATAGTGA	CACACCACTT	GGGCCTATTG	4500
	CTTTGGCACC	AATATGTCCA	ATTCCACCTA	TTACAGCGCC	ACCAATACCA	CCACCAATAC	4560
35	AAGCAGTTAA	GAAAGGTCGA	CCTAATGGCA	AAGTCACACC	ATAGATTAAT	GGTTCTCCGA	4620
	TACCTAGGAA	ACCAACTGGC	AATGCACCTT	TTAAAGTATT	ACGTAATGTT	GTGTTGCGTT	4680
	TACATCTTAC	CCAAAGTGCT	AATGCGGCAC	CTACTTGTC	AGCACCAGCC	ATCGCTGCAA	4740
40	TTGGCAATAA	GTAAGTAGCA	CCTGATTGGT	TAATCATTTT	TATATGAATT	GGCGTAAAAA	4800
	TATGATGAAG	CCCTAACATA	ACTAACGGTA	GGAAGCTTGC	ACCAATGATA	AATCCACTAA	4860
	ATACGCCACC	AATACTAATA	ATTCCGTTAA	CTACTGAAAC	TAAACTGTCT	GAAACAAAAC	4920
45	CTGCTAATGG	CATAAAGATA	AAGATAGTTA	ATAGTCCTAC	AATCAACAAT	GCAATAGTCG	4980
	GCGTTACAAT	AATATCAATC	GCATTTGGCA	CAATTTTATG	TAATCTCTTT	TCGACAATAC	5040
	TTAAAATCCA	AACGGCAAAA	ATAACGCCAA	TAATCCCACC	TTGTCCAGGT	TGCAATGGTT	5100
50	CTCCAGTGAA	GACATTTCATT	AAAATATTTT	TACCAGCAAT	ACCCGTTAAT	AACGTTGTAC	5160
	CACCAATCAC	GCCACCAAGT	CCTGGTGTGC	CACCAAATTC	TTTAGCCGCA	TTAATACCAG	5220
55							

	GCGTAATCCA AGCACCTGAA ATATAGCCTG CCACCATTAA GTTACTCAGT ACTGCTGCAA	5340
	TACCACCAAT TAATCCAGCT CCAATAAATG CAGGAATCAA CGGTATAAAG ATATTGGCAA	5400
5	TTGATTTCAA TACTTTATTC AACTTACCAT TCTTTTGTTC TGCTTTATGC GCTTCCTTAT	5460
	TCGCCCTTTC TTTATCAGCT GCATATGATT TATAGTCCAT TTTTCACTA TCATTGTGAT	5520
	GGTGTGGTAT TGGGTCACCT AGTTTAACAC CACTTAATTC CGCCATATGA TTAGCCACTT	5580
10	TATTGatGTA CCAGGTCCAA CCACAACCTG AATGCGTTCA TCGTGTATAA CACCCATGAC	5640
	ACCATCAATA TGCCTTAGTT CTTGGTCATC TACTTTATTC TCATCTAATA CTTTAATACG	5700
15	CACACGTGTC ATACAGTTCA TGACACTATC TATATTATCC ATACCACCTA CTGCAGCAAT	5760
	AATTCGTTCT GCAAGTTGTT GTTCTTTGGT CATTTAAATC CCTCCTAAGG TTGTCTATCT	5820
	CTGATTGCTC GTTTAAaATG TCACCATTGT TTAATAACCG TCTGTGTGCT TCTTCCTTAG	5880
20	AAATGCCACA CATACCATA ACTGTCGCAA CTTTCACATC ATGCTCAGAT ACCTGATATA	5940
	ACGCCATTGC TTCATCATAT GTGATAGCAC ATATTTCTTG AATAATACGC ACTGAACGGT	6000
	CGATCAGTTT TTGATTGGTT GCTTTAACAT CAATCATGAG GTTATCGTAA ACTTTTCCGA	6060
25	CACCAACCAT TGTGATGGTT GAAATCATAT TTAaAATTAA CTTTGTGct GIACCAGACT	6120
	TTAAACGTGT TGAACCAGTT AATACTTCTG GACCAACTTT AACTTCTACT GGATAcTGCG	6180
	CAATTTCACT TATAACTGCA TGTTCAATTGC ATGAAATAGA TACTGTTGTA GCACCGATTG	6240
30	TGTTAGCAAA TGTTAAACCG CCTATAACAT ATGGCGTTTT GCCACTCGCG GCAATTCCTA	6300
	TAACGACATC TTTTGATGTT AAATCTATAT TTTTCAAATC TTCTTCCGCT AATTTTTTGT	6360
35	GATCTTCCGC ACCTTCTACA GCCATCGTCA TAGCATGTTG TCCACCAGCA ATAATACCTA	6420
	TAATTTCAATG AGGGTCAGTA TTGAATGTAG GTACACACTC CGCTGCATCT AAGACACCCA	6480
	ACCTTCCACT TGTACCTGCA CCGATATAAA TCAATCGTCC ACCCTTTTTTA TACTGTGCAA	6540
40	TTGTTTTTTTT AATTACTTTT GTCAATTGTG GTATTGCCTT TCGAACTGCT AACGGGACTT	6600
	GCTGATCTTC TTTATTCATC GTAATTAAAG CCTCTTCCAC AGTCATTTCA TCAAGATGCA	6660
	TCGTCGCTTC ATTACGCGCT TCGGTCGTAC TATTTTCCAT CACTTCTTAC ACTCCCTAGT	6720
45	TTTTTGAAAA TCAAATGTAT CATTCGGCTC GATACAACTT AACAGTGGTA AGTCTTCTTT	6780
	AATAATTTGT GCAaCAACAT TCACATTGTC ATGTGCACTA AGCGTTTGTC TCACAATTTG	6840
	CATTTGCCTC TGATAACGTC CGTTATTCAA ATTATCAACG GTTACTGAAC CAATGCGTCG	6900
50	TTGCGTCGTA AACTGTGGTT GAATCGAATG TGGACATATT TGTCTTGACG TTTCCGAACG	6960
	AATGACATTT TCCGGATTAT CCGGGCGTAC TTTATGACAC ATATCGAAAA GGTAAGTCAC	7020

55

## EP 0 786 519 A2

	AAGTTGTTTT GCCTGCCTCA TTTCAATCAA TGAGTCTCCA ACTAACACTT CAGATACACC	7140
	AGTTTCTTGT AATAATTTAG CTGCAACGAC AGGATGACTA TGTCTCGTTG CTTCAATTGT	7200
5	TGGCAAGCCT TTATGCAAAG GACCTCGCAA ATCACTCCCT ACAATAAAAC CATATATTTG	7260
	TGCCTTTGGA TTAAATTGAT AAATGAGTTC ATTTTCTTA TTGACCAAGT CAACAGATAA	7320
10	TCCCGTATCT GGTCTTGGAT AATAGTTATG ACAAATGAA AGTAATGTAA AATCATTCAA	7380
	TTGTTGATGT AAGCTTGTTA ACAATTCCCG GGAAATAATA CTTGCATTCA AACAGCACTT	7440
	TAAACCCTGT GCCATTATCG CTTGATTGTC CTCAATTGAT GTACTATGAT CGATACGAAT	7500
15	CATAAATTGT GCATCATATT GTCGAAGATG GTCATAAAAA GATGGTGTTA AAATAGATGG	7560
	ATTAGCATCT ATGAGGTAAG TCACCTGTTC ATGTTTTAAT AAATTGAGTA GTTTTGTGAA	7620
	ATAATGATAT TTTGTCTCGT CATCTTCTTC TGGTATTTGT ACAGATGTAA AAATCATTTG	7680
20	GTAACCTTGT TTAATCATTC GCTTAATATA CGCTTCATCT AAAGGTTGTC CTAAATACAC	7740
	TGAAAAGCCT GTCAAAGTAG CCCTCCTTAA CAATATAATT ATTAGGAAAA TATAGTTGAT	7800
	TTGTGTAATC GCTTACATTT TACTATAAGA GAAAACACAT TACAATATTA ATCAGTTAAA	7860
25	GCCTGTTTCA TGTAATAATC TTACATATTT CTGTCACAAG TTAATTATTA CACCATCAAA	7920
	GATTATCCTT TCTTTTAAAGT GCTGATAATA GCTGCTACTG CTGGATTATT ACAATAACTT	7980
30	TTATACATTT TATTCAGGAT TATCTTATAT TATGTTTTAA TAATAATCTG TGAACAATTA	8040
	AGAGATTTGA AATTGAATTT AATAATTGTA TTGAAAACGC ATACTTCACC ATGCTAAAAT	8100
	AGGAGTCGCA AACAAATAAG ATTCAATAAG ATGTGATGGT TACCAACACA GTCTATTTGC	8160
35	TCGTGTCTTT TTTTATTGAA TCTTAAATAA TAAATACAAC TTTGGAGGTT GGACAAGTGA	8220
	GGAAGAAACT TTTCGGTCAA TTGCAACGTA TTGGTAAAGC GCTAATGTTA CCTGTTGCGA	8280
	TTTTACCAGC AGCTGGTCTG TTATTAGCTA TCGGTACAGC TATGCAAGGT GAATCATTAC	8340
40	AACACTACTT GCCGTTTATA CAAAATGGTG GCGTACAAAC TGTCGCTAAA TTAATGACAG	8400
	GTGCTGGTGG TATCATTTTT GATAACTTGC CTATGATTTT CGCATTAGGT GTCGCAATCG	8460
	GATTAGCTGG CGGTGATGGC GTAGcAGCTA TCGCAGCATT CGTCGGTTAC ATAATCATGA	8520
45	ACAAAACAAT GGGCGACTTT TTACAAGTTA CACCTAAGAA TATTGGTGAT CCAGCGAGTG	8580
	GTTACGCTAG CATTTTAGGT ATCCCAACAT TACAAACAGG TGTGTTCCGC GGTATTATAA	8640
50	TCGGGGCCCT GGCAGCTTGG TGTTATAACA AGTTCTATAA CATTAACTTA CCATCTTATT	8700
	TAGGTTTCTT CGCTGGTAAG CGTTTCGTAC CTATTATGAT GGCTACAACA TCATTATTTT	8760
	TAGCATTTCC AATGGCATT AATTGGCCAA CGATTCAATC AGGATTAAAT GCATTCACTA	8820
55		

EP 0 786 519 A2

	TAITTAATTCC	ATTCCGGTCTA	CATCACATTT	TCCACGCACC	GTTCTGGTTC	GAGTTTGGTT	8940
	CATGGAAAAA	TGCAGCTGGT	GAAATTATTC	ACGGTGACCA	ACGTATCTTT	ATCGAACAAA	9000
5	TTCGTGAAGG	CGCACATTTG	ACAGCTGGTA	AATTCATGCA	AGGTGAATTC	CCTGTTATGA	9060
	TGTTCGGTTT	ACCTGCAGCA	GCTTTAGCAA	TTTATCACAC	AGCTAAACCT	GAAAATAAGA	9120
	AAGTAGTAGC	AGGTTTAATG	GGTCTGCTG	CTTTAACATC	ATTCTTAACT	GGTATTACAG	9180
10	AACCATTAGA	ATTCTCATTC	TTATTTGTAG	CACCATTATT	ATTCTTTATT	CACGCaGTAC	9240
	TTGATGGTTT	ATCATTCTTA	ACATTGTACT	TATTAGATCT	TCATCTAGGT	TATACATTCT	9300
15	CAGGTGGTTT	CATCGACTAC	TTCTTACTCG	GTATACTACC	TAATAAGACA	CAATGGTGGT	9360
	TAGTCATTCC	TGTAGGTCTT	GTATACGCAG	TTATTTACTA	CTTCGTATTC	CGATTCTTAA	9420
	TTGTAAAAAT	AAAATACAAA	ACACCAGGTC	GTGAAGATAA	ACAATCACAA	GCGGCTACTG	9480
20	CTTCAGCAAC	TGAATTACCA	TATGCAGTAT	TAGAAGCTAT	GGGTGGCAAA	GCAAACATTA	9540
	AACATTTAGA	CGCTTGTATC	ACACGTCTAC	GTGTTGAAGT	TAACGACAAA	TCTAAAGTTG	9600
	ATGTTCTTGG	TTTGAAAGAT	TTAGGCGCAT	CTGGTGTATT	AGAAGTCGGC	AATAATATGC	9660
25	AAGCAATTTT	TGGTCCTAAA	TCTGACCAAA	TCAAACATGA	AATGCAACAG	ATTATGAATG	9720
	GTCAAGTAGT	AGAAAATCCT	ACTACTATGG	AAGACGATAA	AGACGAAACT	GTTGTTGTTG	9780
	CAGAAGATAA	ATCTGCAACA	AGCGAATTGA	GCCATATCGT	GCATGCACCA	TTAACTGGTG	9840
30	AAGTAACACC	ATTATCAGAA	GTGCCTGATC	AAGTGTTTCA	CGAAAAAATG	ATGGGTGACG	9900
	GTATCGCTAT	CAAACCTTCA	CAAGGTGAAG	TTCTGTCACC	ATTCAACGGT	AAAGTACAAA	9960
35	TGATTTTCCC	AACAAAACAT	GCAATTGGTC	TTGTATCAGA	TAGTGGTTTA	GAACATTATA	10020
	TCCACATCGG	TTTAGACACT	GTTAAATTAA	ACGGAGAAGG	CTTTACTTTA	CATGTTGAGG	10080
	AAGGTCAAGA	AGTTAAACAA	GGTGATTTAT	TAATCAACTT	TGATTTAGAC	TACATCCGCA	10140
40	ATCATGCAAA	GAGTGATATT	ACGCCTATTA	TCGTGACACA	AGGAAACATT	ACAAACCTTG	10200
	ATTTTAAACA	AGGTGAACAT	GGCAACATTT	CATTGGCGA	TCAATTATTT	GAAGCTAAAT	10260
	AATGCTTACT	ATAAACAGGT	GCGTATACCT	TCATAAGGTG	ACGCGCCTGT	TTTTTCTTTG	10320
45	CTATTGTATT	TTGCAGCATC	ATTGATAGTT	CGCTCTCCCC	TTAAATTTTG	AATTTTAAGA	10380
	TCATCAATTA	AAGCCCCCCT	TCATACTCAT	TTCCTAAAAA	ATATTAATTG	TTCATTATTG	10440
	TTAGCGTTTT	CACAACAAAG	TCAACTTCCT	TGACCTTACA	CTATATTCGA	GGCTATCATT	10500
50	TTAAGTGTA	ATATAGAGAA	AAGGTGGCTT	TTTTTATGAA	ACAACGCATT	GGAGCTTACT	10560
	TAATTGACGC	TATTCATCGA	GCAGGCGTCG	ATAAAATTTT	TGGTGTTTCT	GGTGATTTTA	10620
55							



EP 0 786 519 A2

	ATGAATTAAA CGCAAGTTAC GCAGCGGACG GTTATGCCCG TCTTAATGGA CTCGCTGCAT	10740
	TAGTTACTAC ATTTGGTGTT GGC GAATTAA GTGCCGTCAA CGGTATCGCA GGTTCATATG	10800
5	CTGAACGCAT ACCTGTCATT GCGATTACAG GTGCGCCGAC ACGTGCTGTT GAACAAGGCG	10860
	GTAAATATGT ACATCACTCA CTGGGTGAAG GTACATTTGA CGACTATCGA AAAATGTTTG	10920
	CACATATAAC CGTTGCACAA GGTATATCA CACCTGAAAA TGCAACAACC GAAATACCAC	10980
10	GTTTAATTAA TACAGCAATC GCCGAAAGAC GCCCAGTTCA TTTACATTTA CCAATCGATG	11040
	TCGCAATCTC TGAAATTGAG ATACCGACAC CATTTGAAGT GACGGCAACT AAATATACGG	11100
15	ATGCATCAAC ATATATAGAG TTATTAGCAA CTAAACTGCA TCAAGCGAAG CAGCCTATCA	11160
	TCATTACTGG ACATGAAATT AACAGTTTTT ACCTCCATCA AGAATTAGAA GATTTTGTAA	11220
	ATCAAACACA GATACCAGTA GCACAACTTT CATTAGGAAA AGGTGCTTTT AATGAGGAAA	11280
20	ATCCATATTA TATGGGTATT TACGATGGGA AAATTGCCGA AGATAAAATA CGAGATTATG	11340
	TGGACAACAG CGATTTAATT TTAAATATTG GAGCCAAATT AACAGATTCA GCAACAGCAG	11400
	GTTTTTCATA CCAATTCAAT ATCGATGATG TCGTTATGTT AAATCATCAC AATATCAAAA	11460
25	TTGACGATGT TACAAATGAT GAAATATCTC TACCATCATT GTTAAACAG TTATCCAATA	11520
	TTTCATATAC GAATAACGCA ACGTTCCCTG CGTATCATCG TCCAACATCA CCCGATTATA	11580
	CTGTTGGCAC AGAACCATTA ACACAACAAA CTTATTTTAA AATGATGCAA AATTTCTTAA	11640
30	AACCAAATGA TGTCATCATT GCTGATCAAG GTACATCATT CTTTGGTGCT TATGATTAG	11700
	CATTATACAA AAACAATACT TTTATAGGGC AACC GTTATG GGGTTCTATC GGCTATACAT	11760
35	TACCTGCAAC ATTAGGTTCA CAATTAGCAG ACAAAGATCG TCGTAACTTA TTATTAATTG	11820
	GTGATGGCTC ATTGCAACTA ACTGTTCAAG CTATTTCAAC TATGATTAGA CAGCATATTA	11880
	AACCGTATT ATTTGTGATT AATAATGACG GCTATACGGT AGAACGACTT ATTCACGGCA	11940
40	TGTATGAACC TTATAATGAA ATTCACATGT GGGATTATAA AGCTTTACCA GCTGTATTTG	12000
	GTGGTAAAAA TGTTGAAATT CATGACGTTG AATCATCAAA AGATTTACAA GACACGTTTA	12060
	ATGCAATTAA TGGTCATCCC GATGTGATGC ATTTTGTGCA AGTCAAAATG GCTGTGGAAG	12120
45	aCGCACCGAA GAAACTCATC GATATCGCTA AAGCTTTTTT ACAACAAAAT AAATAATTTC	12180
	ATCGTATACA GGGTATAAGT TTAAGCGAAT ACTTTATTAA ACGAATAGGA CTCTGATATA	12240
	AGATGATTAA TTTTAATAAA ACCGCTTTAG TGTTAATCGA CCTGCAAGAA GGTATTCTTA	12300
50	AAATGGATTA TGCCCCATAT ACAGCTGAAA ATGTCGTTCA AAACGCTAAT AAATTAATAG	12360
	ATGTTTTTAG AAAAAACAAT GGCTTTATCG CTTTGTTCG CGTGAATTTC TATGATGGTA	12420

55

	AGTCGTTTCC ATCATTATT AGACAAGAGA GATGACGATT TTGTCATAGA CAAACGACAT	12540
	TTTAGTGCAAT TTGTAGGAAC AGATTGGGAC TTACAATTGC GACGTCGAGG AATTGATACG	12600
5	ATTGTTCTTG GTGGTGTGCG AACGCATATT GCGGTAGATA CGACAGCGCG AGATGCCTAT	12660
	CAATTAAACT ACAATCAGTT TTTTGTACATA GATATGATGA GTGCACAAAA CGAAACGCTA	12720
	CATCAATTTT CAATAGATAA TGTATTCCCA TTGATGGGAC AAACAATAAC TACAAACGAC	12780
10	TTTCTAAATA TATTGAACTA AACATATACT TCCCCCTTC GATCATGTTG AGGGGGATCT	12840
	TTATTTCACA AAGTATTAAT ACGTCGGGTT GTCTAACCTT CTATATTAA CATATTCTAT	12900
	ATCTGTAAA TCGTTCTTAA CTTACGCCCC TACTACATAA AAAACAGTAT TTATTCCGGA	12960
15	ATTTTCAAAA AATTAGTAT TTATTGCAAA ATTATGTATC ACTTTATGTT TAATTTTGA	13020
	TATTATCTTA ATTAAGTAGA TTTTATAAG TTCTAAAAAG GAGAACAAAT ACATATATGA	13080
	AGAAGAAACT AACATTAAA GAAAACATGT TTATAGGTTT TATGTTATTT GGTTTATTCT	13140
20	TTGGTGCCGG CAATCTTATC TTCCAATAC ACTTGGGTCA AGCTGCTGGT TCTAACGTTT	13200
	TTATCGCTAA CTTAGGATTT TTAATTACAG CAATTGGCTT ACCATTTCTA GGTATCATTG	13260
25	CTATTGGCAT TTCAAAGACA TCTGGTTTAT TTGAAATTGC ATCGCGTGT AATAAACAT	13320
	ATGCTTACAT TTTCACGATT GCCTTATATC TAGTTATCGG ACCATTTTTC GCCTTACCTA	13380
	GACTGGCAAC GACATCATTT GAAATTGCAT TTTGCCATT TTTATCACC AAGCAAATCA	13440
30	CTTTATATTT ATTTATTTT AGCTTCGTCT TCTTTGTGAT TGCATGGTTT TTTGCGAGAA	13500
	AGCCATCAAG AATTTTAGAA TATATCGGTA AATTTTAAA TCCGGTATTC TTAGTATTAT	13560
	TAGCAATTAT TTTATTATTT GCTTTTATCC ATCCATTAGG TGGCATATCT GATGCACCTA	13620
35	TTAGTAAACA ATATCAATCA CATGCCTTAT TTAACGGCTT TTTAGATGGA TACAATACCT	13680
	TAGATGCGCT AGCGTCATTG GCATTTGGTA TTATCATTTG TGCAACGATT AAAAAGTTAG	13740
	GTATCGAAAA TCCAACGAT ATCGCTAAAG AAACAATTAA GTCTGGTACT ATCAGTATCA	13800
40	TTATGATGGG GATCATTTAT ACCCTACTAG CAATCATGGG TACATTAAGT ATTGGTCATT	13860
	TCAAACCTAG TGAAATGGT GGTATTGCCT TAGCGCAAAT TACTCAATAC TACTTAGGTA	13920
45	ACTACGGTAT CGTCCTGTTG TCACTTATCG TTATGGTTGC TTGTTTAAAA ACAGCCATCG	13980
	GTTTGATTAC GGCAATTTCA GAAACATTCG AACACCTTTT CCCTAAAATG AATTACCTAG	14040
	CGATTGCAAC AGTTGTAAGC TTTATTTTCT TCTTATTCGC GAATGTTGGT TTAACATAAG	14100
50	TTATTATGTA CTCAGTCCCA GTGTTAATGT TCTTATATCC ATTAGCAATT GCCTTGATTG	14160
	TACTAACATT ATTTAGTAGC AAATTCATC ATTCAAACT TATTTATCAA TGTACCATT	14220
55		

	GCACATCATT CTCACAAACT TTGATTAATT TCAGCCAAAA ATATTTACCA TTATCAGACA	14340
	TTGGTATGGG CTGGGTGTT CTCAGTTTGA TTGGTTTCAT TATCGGCTTC ATTATTTATA	14400
5	AAATTAAGCA TCGTAAAT CCACAAGCAT AATACTATGC CACAGTCATA TGTAAACAT	14460
	ATGCTTGTGG CATTTTTTAT TCATACTACA TTAAACTGCA ATCGTATACA TACATATCAA	14520
	TGATTATCCA CAAAAATAT TAGTACTTTC ATTTTACAAA TCACATTAAT ACAAACACAA	14580
10	CCTTATCTTT ATATTATTAA ATTTATATTT GACACTTATA TTGAACAACT GTAATATATT	14640
	AATATTAATT CTTTAAATG TATAAATATA AAGGAGGGAG ACCGATGaAT TCAATCATTG	14700
15	aATTAAGTGA TTATTATAGC TCTAATAATT ATGCACCACT TAAGCTTGTC ATTTCTAAAG	14760
	GTAAAGGTGT CAAAGTTTGG GATACTGATG GCAAACAATA TATAGATTGC ATTTCTGGGTT	14820
	TTTCAGTTGC AAACCAAGGC CATTGTCATC CAACAATTGT TAAAGCGATG ACAGAACAAG	14880
20	CTTCAAAGTT GTCTATCATT TCACGTGTCC TTTATAGTGA CAATCTCGGG AAATGGGAAG	14940
	AAAAAATTTG TCATCTTGCT AAGAAAGACA AAGTACTCCC CCTTAACTCT GGTACTGAAG	15000
	CTGTTGAAGC AGCCATTAAA ATTGCTAGAA AATGGGGCTC TGAAGTTAAA GGCATTACTG	15060
25	ACGGACAAGT TGAAATCATC GCTATGAATA ACAATTTTCA CGGTCGTACA CTTGGCTCAT	15120
	TATCACTATC TAACCACGAC GCATATAAAG CAGGATTTCA CCCCCTACTT CAAGGCACTA	15180
	CAACAGTAGA TTTTGGAGAC ATTGAACAAT TAACACAAGC TATTTACCG AATACAGCAG	15240
30	CAATTATTTT GGAACCAATT CAAGGTGAAG GTGGCGTTAA TATACCACCG AAAGGATATA	15300
	TTCAAGCTGT GCGTCAACTA TGTGATAAAC ATCAAATATT ATTGATTGCA GATGAAATTC	15360
	AAGTTGGTCT TGGTAGAACT GGGAAATGGT TTGCTATGGa ATGGGAGCAA GTCGTTCCAG	15420
35	ACATTTATAT TTTAGGTAAG GCATTGGGTG GCGGCTTATA CCCTGTATCT GCTGTACTTG	15480
	CAAATAATGA TGTCATGCGT GTTCTAACAC CAGGTACACA TGGTTCAACA TTTGGTGGTA	15540
40	ACCCTTTAGC CATTGCAATA TCGACGGCAG CGCTTGATGT ACTTAAAGAT GAACAACTGG	15600
	TTGAACGATC AGAACGCTTA GGTTCAATTTT TATTTAAAGC GTTGCTACAA CTTAAACATC	15660
	CTAGTATTAA AGAAATTAGA GGTCGTGGTT TATTTATAGG CATAGAGCTT AACACAGATG	15720
45	CTGCACCTTT TGTGGATCAA CTGATTCAAC GTGGAATCTT ATGCAAAGAC ACGCATCGTA	15780
	CTATCATTCG ATTGTCTCCA CCTCTAGTCA TTGATAAAGA GGAAATCCAT CAAATTGTTG	15840
	CAGCTTTTCA AGACGTTTTT AAAAATTAAC AATTAATCAT TTATATATGA CATAGGAGGG	15900
50	ATTCATGATG ATTAAAGTAG GTATCGTTGG CGGTAgCGGT TATGGCGCAA TTGAATTAAT	15960
	TCGATTGTTA CAAACACATC CTCATGTAAC GATTGCACAC ATCTACTCAC ATTCAAAAGT	16020

55

ACTTACAGTG GaTAATAATG ACTGTGATGT AATTTTCTTT GCGACACCAG CACCCGTAAG 16140  
 TAAAACATGT ATCCCTCCCT TAGTAGAAAA AGGTATTCAT GTTATCGATT TATCTGGCGC 16200  
 5 ATTTAGAATT AAGAATCGTG AAATATATGA AGCATATTAC AAAGAACTG CTGCAGCACA 16260  
 AGATGATTTG AATCATGCTA TTTACAGCAT TTCAGAATGG CAATCGTTTG ATAACAATGG 16320  
 AACGAAGCTC ATTTCTAATC CTGGCTGTTT CCCTACAGCA ACATTATTAG CATTACATCC 16380  
 10 ACTTATTAGC GAAAAAATAG TAGATTTGTC ATCTATTATT ATTGATGCTA AGACCGGCGT 16440  
 GTCAGGTGCT GGTCTTCAT TATCACAACG gTTcATTTTT CAGAAATGAA TGAAAATCTA 16500  
 15 AGCGCTTATG CAATCGGAAA CCATAAACAC AAACCGGAAA TCGAGCAATA TTTATCTATC 16560  
 ATTGCGGGTC AAGATGTATC AGTCATATTT ACACCACATC TCGTACCAAT GACACGAGGT 16620  
 ATTTTATCAA CAATATATGT CAAATTATCA TCTGAATATA CGACTGAATC ATTACATAAA 16680  
 20 TTAATGACCT CTTATTATGC TAATCAGCCA TTTGTCAGAA TTAGAGATAT TGGGACTTTT 16740  
 CCAACCACAA AAGAAGTACT CGGTAGTAAC TACTGCGATA TCGGCATCTA TGTAGATGAA 16800  
 ACAACGCAAA CAGCAATTTT AGTATCAGTG ATTGATAACC TTGTCAAAGG CGCAAGTGGG 16860  
 25 CAAGCCATTC AAAATTAA AATATTATAT GATTTTGAAG TGACGACTGG CCTAAATCAA 16920  
 TCACCAGTTT ATCCATAAGG GGTGTTAGAA TGAAACATCA AGAAACGACA TCACAACAAT 16980  
 ATAACTTTTC AATTATTAAA CATGGCGATA TCAGTACACC TCAAGGCTTC ACGGCTGGTG 17040  
 30 GTATGCACAT CGGTTTACGC GCTAACAAAA AAGACTTTGG GTGGATTTAC TCATCGTCTT 17100  
 TGGCAAGTGC AGCTGCCGTA TATACTTTAA ATCAGTTTAA AGCTGCACCA CTTATTGTCA 17160  
 CTGAAGACAC TTTACAAAAG TCTAAAGGAA AATTACAAGC ACTTGTTGTT AATTCAGCTA 17220  
 35 ATGCAAATTC TTGTACCGGT CAACAAGGCA TAGATGATGC ACGACAAACA CAAACATGGG 17280  
 TTGCTCAACA ACTTCAAATA CCATCTGAGC ATGTTGCTGT TGCTTCAACT GGGGTCATTG 17340  
 GTGAATATTT GCCTATGGAT AAAATTAAGA CTGGGACCGA ACATATTAAG GATGCTAATT 17400  
 40 TTGCAACGCC AGGTGCGTTT AACGAGGCAA TTTTAACAAC TGATACCTGT ACAAACATA 17460  
 TCGCTGTATC ACTAAAAATC GATGGTAAAA CCGTTACAAT TGGTGGTAGC ACCAAAGGTT 17520  
 45 CAGGTATGAT TCACCCAAAT ATGGCTACCA TGCTTGCTTT TATAACAACC GATGCATCGA 17580  
 TTGAATCGAA TACACTTCAT CAATTATTAA AATCTTCGAC TGACCATACA TTTAATATGA 17640  
 TTAATGTTGA TGGCGATACA AGTACAAATG ACATGGTATT AGTCATGGCA AATCACCAAG 17700  
 50 TTGAACACCA AATACTTAGT CAAGACCATC CACAATGGGA AACATTTGTT GATGCATTCA 17760  
 ATTTTGTCTG TACATTTTTA GCTAAAGCTA TAGCCAGAGA TGGCGAAGGC GCAACAAAGT 17820

55

	CTATCGTAAG TTCAAATCTA GTAAAATCAG CTATTTTTGG CGAAGATGCC AATTTTGGTC	17940
	GAATCATTAC AGCTATTGGC TACAGCGGAT GTGAAATTGA TCCTAACTGC ACATATGTTC	18000
5	AACTGAACCA AATACCTGTC GTTGATAAAG GTATGGCTGT ACTATTTGAT GAGCAAGCTA	18060
	TGTCGAATAC ATTAACATCAT GAAAATGTCA CAATTGACGT TCAGCTTGGT TTAGGTAACG	18120
	CTGCAGCGAC TGCATACGGT TGTGATTTAT CCTATGATTA TGTGCGTATC AACGCATCAT	18180
10	ATCGAACATA AGGTGGTGTT GGTTAGATGA AATTTATTGT CATTAAAAAT GGTGGCAGTA	18240
	CACTTAGTGA CATGCATCCA TCAATTATTA ACAACATTAA GCATTTACGA TCAAACAACA	18300
15	TCTACCCCAT TATCGTTCAT GCGGTTGGCC CATTIATTAA TGAAGCATT TCAAACCAGC	18360
	AAATCGAGCC ACACTTTGTT AATGGCCTAA GAGTGACTGA TAAAGCAACC ATGACCATT	18420
	CTAAACACAC GCTCATTGCA GACGTTAACA CTGCATTAGT AGCTCAATTT AACCAGCACC	18480
20	AATGTTCTGC AATAGGCTTA TGTGGTTTGG ATGCACAGCT GTTTGAAATT ACATCTTTTG	18540
	ATCAACAATA TGGATATGTC GGTGTTCCGA CCGCTTTAAA TAAGGATGCT TTACAGTATT	18600
	TATGTACTAA ATTTGTACCT ATCATCAATT CGATTGGTTT CAATAACCAT GATGGAGAAT	18660
25	TTTACAATAT TAATGCTGAC ACGCTTGCCT ATTTTATTGC ATCATCATT AAAGCGCCTA	18720
	TTTATGTATT AAGTAATATT GCAGGTGTAC TCATCAATGA TGTGTTATA CCTCAATTGC	18780
	CATTAGTCGA TATTCATCAA TATATTGAAC ATGGTGATAT TTATGGAGGT ATGATTCCCA	18840
30	AAGTGCTAGA TGCCAAAAAT GCGATTGAAA ATGGCTGTCC TAAAGTTATC ATTGCATCAG	18900
	GAAACAAGCC AAATATCATT GAATCTATTT ACAATAATGA TTTTGTGGC ACAACAATCC	18960
	TTAATTCATA ACTATGAAAT TAAGGCCTAA CAAGTTTGA CACGCGAGAT GATTCCAGTT	19020
35	CGATTATCCA TTGCGCTAAA ACATTTATTT ACCGTTTCATC TCGTTAACAA TTTGAATAC	19080
	AGTACGATAC AATATGAGAT GTAAAAAACT AATAACCTTT TACAAATTG TTTATCAAAA	19140
40	TATTTTAAGT TTTGCAAAGC TTTTATTGT GATTATTTTC ACAAATACT ATAATGAGGA	19200
	TAGTAAATAG AGAGGAGTCC TTAAGTTGAC GAAACGACAA ATGGGTATAT TCATTTATGC	19260
	TGGAATTATC GGTGGCTTGT TATCTGGAAT TGTAATAATTA GGTTGGGAGG TCATGTTTCC	19320
45	ACCTCGCACA CCAGAACGTA ATGCAACGAA CCCACCTCAA GAGTTATTGC AACCAATTAGG	19380
	ATTTAGTAGT GAGTTTACGC ATCAAACATA TACATTTTCA AATATGGAAT TGCCTTGGGT	19440
	AAGCTTTATT GTCCACTTTA GTTTTTCTAT CGTCATTGCA ATTATTTACT GCATATTAGT	19500
50	TAAAAAATAC GCTTACTTAG CAATGGGACA AGGTGCTGTT TTTGGTATTG CTATTTGGGT	19560
	ATTATTCCAC CTTATCATT TGCCAATCAT GCATACTGTA CTGCTGTGT GGGATCAACC	19620
55		

AGTGCACAA CATTTTGTCT ATCGCTATAA ATTAAATTAA TACTGACT AACATTAACG 19740  
 TGAGTTTtAA ATCATCGTTT GAGTATGATG ATTGATGCTC ACGTTATTTT ATTAAGTAC 19800  
 5 ATGATATGAT TCCAGCCAAC TTACGTGAGC ATTAAAGTCT CAAATGCGTC GTAACAAACT 19860  
 ATTATTTTCG GTAATTTCAA TATTGCTCAG TATATTTTTA CCTTATCACT TACTTTAATC 19920  
 TCGTCATGAT TTTGAATGAT GCCATCGTGT ATTCACCTTT CATTTTTCCA ATAAAAAAC 19980  
 10 ATCTAACAGT AAACATTTAG GCAGTATAGT TTAAAATCAC TCGCAATGA TACTGTCAGA 20040  
 CGTCATATTA ACTACTCAAT AACTGAAATA CAGACACTTT TTTATAACCC CAGGGTGCCT 20100  
 GTCCTAAGAA ACATACCTGT ACCATAAACT GATCAAAAAT AAATTGTTTG AACTTCACTT 20160  
 15 CACGTGATTG ATAAAAGTGT GATTGTGTCA TATCATAAAT GTCCAATCCT TTGATTAAAC 20220  
 CTTACCAAT CAATTTTGTA AAACCTTTCTT TTTGTGTCCA TATTTGATAA AAATCATTTA 20280  
 20 AACTACATAT TTGATGTGCT TCGTTTGTAG AGAAACACGT CACTAACGTA CGCCAGTCTA 20340  
 AACGTTGTGA TATCTTTTCG ATATCAATAC CAACTGGTTC TTTATCGACA ACACACACGA 20400  
 TATAAGGATA ACTATATGAT AAGCTCACAT AGATGGGCTG TCCATCACGA TTGTGTGAA 20460  
 25 CAATATCTGC CTTACCTCGT GCGGAAATGT GATAATGCCA TTCATGTGGT AATAAACCTG 20520  
 TGTGATGTTG AATTCCATAT TGCACTAAAA TATCTCCCAA TCTGTGCATG AGTTTATCTT 20580  
 GATTGTATCT ATAGTTGACT GTACGCGGTT TTTTATATGA CCAACGACTT TGTGATATTA 20640  
 30 ATTCTTCAAT ACTTTTCAAG TTACTCTGTA ATTGCATTAC AAATACTGTC ATAACCTTCC 20700  
 CTACTTACTT ATTGAATATT GTTTTGATAT ATTGTGCCCA ATGATACAGC CAATTGTTAG 20760  
 TTATCGTTGG CCATTTTCA CTGATGTGAT TCATTATTTT TAATGTTAAT GTTGATCTA 20820  
 35 TCATTGCTAG TTGTTGTTCA CGGTCAACAC TAGTTAATCC AATCGTTTCG TACATGTCTT 20880  
 GTTCTGTAA AATTTTCAAT AATGATTCAT CGCTGACGAG TTCAATTTCT TTGCGCTTAA 20940  
 CGCATTCTAA CAAAGATTTC ACCGGCATT TATTAGGTGA TAGCACATGG TAAATGATTT 21000  
 40 GTGGTGTGTT GACCTGTGCT AATGCGACAA TTTGTCTTGC AGTCGTATCC ACAAAGAAA 21060  
 AATCTACAGG CATTTACAGC ATGCTAACCC CGATACAATC CAGTTGTAAC AAATCATTC 21120  
 TTACCATGTA AAAACGGTTA GTCTTTATAT TTCTCATATG CCATCTTCCA TTGTAAGGAT 21180  
 TCGTCAAATT ACCAACACGT ACAATCCGAC CATCTAAGCC ATTATTTACA GCTTCTAATA 21240  
 CTTTAAATTC ACTATAAAAT TTGCTCCGTG TATATGGTGA TGTTAGTAGT TGCCCTTTAT 21300  
 50 AGACATCCGC TTsTGAAAAT GTCACATCTT CTGTGTCTAT ATCAAAATAA GTTCCACAC 21360  
 TTATCGTAGA CACATATATT AACCTTGCAT GATGTTGTTG TGCCAAACGT ATGACATCAA 21420

55

CACCTGCATG AATAATCGTA TCCATGTTTT CTGGTAAAAC AACATCATCC ATACACTCGA 21540  
AATCACCAAC AATGACTTCA ATGTTTGATA ACATTATTTT AACCGTCTCT TCTGAAAAAT 21600  
5 AATCATTTAA ATTCGTCATC AACTTATACC ATGCTATTTT CTCATTATCA GCACGTATGA 21660  
AACAATAAAT GCGATGACTG TATCCTTGTA GTACTTCAAT CAGATAAGCA CCTAAAAAAC 21720  
CTGTCGCGCC AGTCAATAGT GTATTTCCCTA GAGGTCGATG ACTTAGACTA TCCTCTAAAA 21780  
10 TACCCAAGTT ATAACGAGAC ATAACAATCT TTTGTAATTC CGAAAGATTA TCCGGTAATG 21840  
CAACTAATGA TTGTTGATTT TGGTACATAT AATTAACAAT CTGTCGCACG GTTTTATATT 21900  
GGTATAATGT CTGCATTGAA ATATGATGGC CAAATCGTTT TAAATGCGAG ACAACTAACA 21960  
15 TCGCCTCTAA TGAGTTACCA CCAAGTTCAA AGAAATCATC GTCAACACCG ACATCATTTT 22020  
GTTTCAATAC CTCTCCAAAT ACATCAACAA ATGTCTGCTC AATTTTATTA GAGGGTTCGC 22080  
TATACACTTT ATTAGACTGT TGTATAGGTG ATGGATTGGG CAAACGCGTA GTATCCACCT 22140  
20 TGTCATTTCG GTTAATGGC ATACAATCGA TATGCGTTAT AGTCTTAGGA ATCATATACT 22200  
TAGGCAGCTG ATCATTTAAA TATTGCTTCA AATCCTGTTT CACTTGTTGC TCTCCGACAT 22260  
25 AATAAGCATT CAATATATCA TCGGTATCAA AGTGAATTAC TGTTACAACA CAATCAGATA 22320  
TACCACGAAT AGCTAATATT GCATTTTCAA TTTTCATCAAG TTCAATACGG TACCCGTTAA 22380  
CTTTCACTTG TTTATCTATT CTTCTAAAA ATTCAATTG ACCATCAGAT GTATAACGTG 22440  
30 CTAAATCACC ACTATGATAC AACTTTCCCT TACCAAATGG ATTATTTTGC CATTTATCAG 22500  
CCATTAATTC TGGACGATTA ATATATCCTA TCGCTAACT ATCACCTGCA ATACACAAC 22560  
CGCCTGGCAT ACCAATACCG CATAACAAAC CATCTGACAT AATATACACT TGGATGTTAG 22620  
35 ATAAGGGTTT GCCAATTGGA ATCGTCTCAG GTATCAAATC ACCACAATGA TGTGACCAAT 22680  
ACGATGTGAT GACTGTTGAC TCAGATGGTC CATAGGCATT GAAATACGTG CCACAATGCT 22740  
TCTCAATATA TTTAACAAAG GATGCCGTAC TAGTTGCCCC GCCTGTAATC AACTTTTCAA 22800  
40 TATAAAAGTC TTCCATAACA CTACACATCT GTAACGGAAT CGACGCAACC GTCACACGAT 22860  
GCTTATTAAT GAGTTGTTGT AACTGTTCTG GATTAACACG TTCCTCTCTA TCTGGAATCA 22920  
CAAGCGTATG ACCATTTAAC AAACAACAAT AAATCTCCAT AACTGATGCA TCAAAAACAA 22980  
TATTTGCATG TTGCAAAAAT ACTTCATTGT CGCCTAATTG CAATTCAGTT GACCATGCAT 23040  
GCACTAAATT CAACAAATTT CGTTGTCGTA TGGCAACCCC TTTAGGCATC CCGGTCGTAC 23100  
50 CAGATGTGTA AATAGCATAC ATCTCATTAT CTAACATCGC TGTGTTTTCA AGTTGATTGC 23160  
CATGTAAATC ATCATATTGT TCATTTTCCT TTGATTCAAC AAAGCCTTTA GCATTTTCCA 23220

55

	TAGCATCCTC	CAAAATTGCA	CCTTGTCGTT	TATTCGGAAA	ATCAATATCG	ATAGGTATAT	23340
	AAGATGCACC	TACTTTAACT	GTCGCCAACA	TCGCCGCAAT	CATTTCAAAA	CTACGTTCTG	23400
5	TAAACAAGGC	AACCCGTTGA	CCATTGCCCA	CACCATTGA	TAGGAGCATG	TGCGCAATGG	23460
	CATCCACATA	GTTGCGTAAT	GTTTCATACG	TCATTGTCAA	ATCATTCATG	ACTAGCGCAA	23520
	CATGATTACC	TTGTCGTGAG	ACAACCTTCAT	TAAAGTAACT	TATGATAGAT	TTATTTCCTG	23580
10	GGACATTAAG	CATTCGATCG	TTAACATGCG	TATTGACCCA	ATTTAGAAGT	TCCTCCGTGC	23640
	CGTTTGGTAT	ATCACAAATT	TGTAGTGTAT	CTTGATGCTT	CAAAATATAA	TCAATCATAA	23700
	TCATACATTG	ATTACCCATG	TGACGAACTG	TTTCTGAGTG	ATATAAATCG	GTATTATACT	23760
15	CGATATTGAT	TGTATAGTCA	TCGCGATCTT	CTTCAATGAT	GAAAGATAAA	TCAAATTTCTG	23820
	CCGTCACTGA	TTTGGGTTGA	ATGTGTGTTA	ATTTACTATG	CCCAAAATGA	GCATGATTCTG	23880
20	TTTCATGTGT	TTGTAGTACT	AACATGACAT	CAAATAATGG	ATTCCGTGAG	GCATCATGTG	23940
	ATTGATCTAA	GTCAATTTACT	AAACATTCTGA	ATGGGTATTC	TTGATGCTCG	TATGCCTCCA	24000
	AACTCATTTT	CTTAACCTCT	TGTAAAACT	GTGTCCACAT	TTTATCAGGT	GACGGTTGCC	24060
25	CTCTATATAC	CAACGTATTA	GCAAACATGC	CTAGCATTTG	CTCCGTGCCT	TTATGCATAC	24120
	GCGCACTCAT	CACACTACCG	ACAACAACAT	CATCTTTTCG	AGCATATCTA	CTTAACAACG	24180
	TCATGACCAC	ACTCATAAAG	AACATAAAAT	CAGTAATTTG	ATGCTTTTCT	ACATACTTTT	24240
30	GAAGTAGCTG	TCTCATTTGT	TGATTCATTG	TAAATGACAT	CATTGCTCCA	TTTGTGTTTT	24300
	TAATATTTGG	TCTAACATAG	TCTGTCGGTA	AGCTTAAAAT	AGGTACTTCA	TCTTTGAATT	24360
	GAGATAACCA	ATATTGTCTA	TGTTTCGTCA	TATCAGCATG	CGACATCCAC	TCACTATAGT	24420
35	CTTTATATTG	CAATTTAAGT	GGTAACAATA	ATTTATGTTG	ATAAAGTGCG	TTAAGATCAT	24480
	TCAATTAATTG	TATATTACTC	ATACCGTCAT	TAATGATATG	ATGCGTATCT	ATAAAGAGGT	24540
	ATGCATGTAA	GGGACTTCTA	ATGTATCTCA	CTCTAATTTG	ACTTGGCTTT	TCCAAATTAA	24600
40	AAGGTGCTAC	AAATTGGCGC	ATGATTTCTT	GTTTCATCCG	AAAATGCGTG	TTAATTCTTT	24660
	CAAAGTCAAC	TGCAACATCT	GCCACAATAC	GTTGTGGAAC	CTCATCATCT	ACAACAATAT	24720
45	ATTGTGTTCG	TAAAATCTCA	TGTGCGGCTA	TCAAACGCTG	CACTGCTTGT	CGCAATTGAG	24780
	CTACATTAAG	TTCTGATGAT	AACCGCCATA	AAAAAGGTAC	GTTATACACC	GTATCTTTAT	24840
	GGTTTGATTT	CCATAATAAA	TACATACGCT	TTTGTGCAGA	GCTCAGCACA	TAATCATCTT	24900
50	TAACATATAGT	TTCTGGAATC	ACTTCATAGT	TTTGTTCCTG	AACCTTAGCA	ATCGCTTGTG	24960
	CTAGTTCAAA	TACAGTTGGC	TTTGTGAATA	AATCACCAAT	TTGTAATCGT	TTCCAGTAG	25020

55



	AATTATCATG AATACCTACT TGATTCACAT GTAAAATATC TGCAAAAATT TGGCATAGCA	25140
	AGTGTTCCGT ATCTGTACTC GGTGCTACAT AGGCATCCGT ATCGACATAG TCCATGATAG	25200
5	GCAATGCCTT CTTATCTAAT TTCCCATTA TAGTAATAGG AATTTGCTCA ATATGCATGA	25260
	AATTAAC TGG TATCATGTAC TCCGGTAAGG TCATACGTAA TTGTGATTTA ATCTTATTAT	25320
	GTGATAATGT ATGCATCGCT TCATAATAAG CAACGATATA CTGATCTTGA TCATGATTTT	25380
10	GAACAATAAC AACTGCTTTA TTAATACCTT GTATACGCTC GAGCGCATGC TCAACCTCTG	25440
	ACAAC TCAAT CCTAAACCCT CGAATCTTAA CTGTGTTGTC CTTTCGATAT AAATAATCTA	25500
15	TGTTGCCATC GGGTAACAAA CGAACGATAT CACCACTTCT ATACATCAGC TGaTTTATAT	25560
	TTGAATCTTT GATAAATTTA TCTGCTGTCA ATTCTGGCTG ATTTAAATAA CCTGCAGCTA	25620
	ACCCAAAGCC ACTTG TACAT AATTCTCCAG GAATACCAAC GCCACACCGA CGCTCGCCTT	25680
20	GCATGATATA AACATGAGTA CCCAGAATCG GTTTACCAAT AGGAATACGA TTTGGAAC TT	25740
	TGTTAGGTAT ATTATACGTC GTTGTAATG TTGTATTTTC AGTTGGTCCA TAACCATTAA	25800
	TAATTTGAGG ATGCTTCGGT TTTTGATTAA GCAAATCCAC CCACTTAGCA TTCAATACTT	25860
25	CTCCACCAAT TAATAAATAC TTTAACGGTA CCAATACTTC TATTCGTTCA CTAGCAATCT	25920
	GATTAAATAA TGAGGAGGTT AACCACATAG TATTAACGTC ATTTTCATTG ATTAATTGTT	25980
	CTACCGCTAT TGGATTTAAT AATTGTTCTT TTTTAGCAAC AATCAGCTTT CCACCATTGA	26040
30	GCAATGCACC ATATATTTCA AATGTTGCAG CATCAAAGGC TATAGTTCCT GATAACAAAA	26100
	TCGTCGCTC TTCATTTAAT GGTACATAAT GATTTTGATG GACCAAGCGA ACAATACCTC	26160
	GGTGCGGAAT TAGTGTCCCT TTAGGGTTAC CAGTTGTCCC CGACGTGTAA ATAACATAAG	26220
35	CATGATCTTC TAACGTGTTA CATTTAGAAA GATTATCAAT ATTTTTCAC GCTATCTTAT	26280
	TCAAATCAAT GTGATTAATA TTTTGTTTAC CATTTTCATA TAAAGCTTGG TACGTTATTA	26340
40	CAACTTTAGG CGTTACATCT TTTAAAATGT ACTCCTGACG ATCACTTGGA TAGTTCGGAT	26400
	CAATTGGCAC GTAAGCCCCA CCAGCTTTCA ACACACCTAT CATCGCTATT ATCATCTCAA	26460
	TACTTTTTTC AGCTATGACA GCGACACGAT CATTAGGTTT AACACCATAC TGGTTTCTCA	26520
45	AACGGTGTGC TAAATCATT CCGCGTGCAT TCAATGTTTG ATATGTTATA AACaCTCCGT	26580
	CAAATTGCAC AGCGACATGA TTCGGCGTTG CTTCAACTTG TTGCTCAAAT AAGGTAACAA	26640
	CTGTTTGCGC ATCATCTATC TCAGGCAAAC TTAAATTGAT ATCGTCATAT AATTGAATAT	26700
50	CACGTTCTGT CATCAAATTA AGTTCATCTA CAGTTGTTG TTTATTTCCA TTTTCTTCAG	26760
	TAATTTGCAA ATAAATATTT CGAACTAAGT CACTCAGCGT CTCGATTGAG AGCAAATCAT	26820

55

	CTAAAGATGT ATGTGCATCA TGTATTTGAT GTACATCCTC AATAACATCA TTACAACAAG	26940
5	ACATCATATG ATGATAACAA TGAAAAATAG TCTCTAGTGA AAGCGAAGAC TTCGCGCACT	27000
	GTAATTGCGA CATATTTTGC AACACACATT TATTAAAAATC TGTGTAAAA CGTTGACATA	27060
	CATCTTTTGC ATCGATTGTT AACGTTAACG GCACAATATT TCCGTGTAAA TCATTTGGTA	27120
10	AATGTGATGG TACATGTATA CCTAATGTGA CATCATGTTG TTGACTCATT ATATGATTAG	27180
	CTAAATACAC ACTAACAGCC AACGATGCCA TATCTATTGA TGTCATATCA TCAATCAAAT	27240
	ACGTTTGATA TAAAGCTTGT TCAAATGGAT GCTTAATTGG AAAATAACTA TCAATATGGA	27300
15	TGTCAGAGTT ATTCTCTAAC CGAAAATAGT TTGAGTCTAA TGCTATATGC GATGCATCTT	27360
	GATTGTCTTT ATCATCATTT CTATTTATAT GTGCATGCTG TCGAGTATTG TTAATAACAG	27420
	TATTGCCACG ATATGCATTG CATAAATCAT CAAGAAaAAT ATCAATTTGA CTATCATCga	27480
20	AAATGGaCAC ATGAAAATCT aATAGTATAT ATGcAGCATC AGCGAACTGm AACAAATTTAA	27540
	CTTTGAATAA AGGTGAATCA TTAAAATGGT AAGTACTTAA TTCTTGCTTA AAAAAAGCTT	27600
	CTAAATCATA GTTTGCGGAA GAAGATGGAA CTTGTTTTAT CTCAATAAAA GGCAGAAATT	27660
25	CATGAAGTAT CATTTGTTAA TTGTCATCGG TAGTAACATC AAAAAAATGT CTTATAGATG	27720
	CATGTTGTgC ACAATTGTCG ATAATGCATA CATCATTTTA GTAGCTTCAA CATTTTtagC	27780
	GAGTTTAACC CAATACGCAT TACGGTGTGT CGTTGATTCT GTATTATTTT TGTATATACG	27840
30	AAAATATTCC TGTTGAAATC TCAAATTACC CATAATCATA AAAAGTCCTT CTTTCATATC	27900
	ATAATACTCA TTACTTACTG AAATTGCATG ATGATATGAT AACCGACGAA ATGTTAATTA	27960
35	ACTCGTTATG TAATGaTTAA TATAAAACAC CATTGCAAC ATATGAGCGA TATATTCTAC	28020
	CCTAAAATAC ATCTTGTATC ATCGTTACAA TTGGTATATT TTTCAATGTA AATTACATAC	28080
	ATCTTCGATA AATAGCACAC TACAAATCGT TAATCACTTT CTGTTGTTCa CATCTCATTG	28140
40	CAAACTCAAT ATTGTTGTTA CAAAATATCC ATGAAGCAAG TTTATATTAA ACAAACAAC	28200
	CGCATAAAAC AATTGTTATC CTTAAATTTT AACAAATTCT TAATAAATTT ATCTCTATTT	28260
	TAATTACGAC CAAATTAATA GGTTTTCCAT ATAAAAAGAT GCATAAAATA AATATTTAAA	28320
45	TAAATTCAAT TTGTATTAC TTGTTTTGT CCCCCAAATA CACCAGCAAC AAGCATGCTA	28380
	GCACCAATTG TTAAAACGAT AAACATATAC AGTCCCATTT GTAATGACGT TAAGAAAACA	28440
	CCCAACACAA TCCCTAACCT AGCTAGTGTT TCTGAAAAAT GAATACCTAA TGCATTAACT	28500
50	GCACTATATG TTCCTCTTTT AGCTTTAGGA ATAATTTTAA AGCGTTGTTC TGAAACTATA	28560
	GGCGAATAAA TAATTTcACC TACAGTCGCA ATTATCATAA AAACAACCTAA TAAGCCAAAC	28620

55

GCTTTTTTAA AATCTATTTT CAATACAACT TTCGAGATTG AATACGTGAG TAaAATGACG 28740  
 ACGACCGTAT TAATCATTAG CAAGATTGCT AACATCTTAG CACCTGTAAT ATCATATGAA 28800  
 5 CCTATACTTA TTGTTTCAAA CTGATCCTTT AGTCTAATAG CAATATATGA GGAGATTGAA 28860  
 AATTCACCCA TCATGATGAT ACTGAACCCC GAAATCAATA ACATATAATT ACGGTCTTTC 28920  
 AAAACTAATT TATAACTGCG AAATATATTC ATTATTTGTA ATTTTGTGATA ACGACTTGCA 28980  
 10 TGCCTCTTGT CATCACTTTG CTTTACTTGA TTTTCGGTCTT GAGGTAACCA AATATATAAA 29040  
 ATAAAGAGTA CAATTAAAAA TATACAAGCT GCTATTAAGA AAAGTAGTAA CATACTGTAG 29100  
 CCATACATCA AGCCACCTAA CAATGCCCCA ATAGCTACCG ATAAGTTTGT CATCCAATAG 29160  
 15 CTAATCTTGT AAATATAATG TTCCACGTCT TCGGTAATTG CATCCATAAT TAATGTGTCC 29220  
 ATAAGTGGAA ATTGTAATCC CCAAACGATT GTAAATATGG CATATGCAAC AAAAAACCA 29280  
 ATAATTTGCC ACAATTGATG TGACCCAAAT ACGCCCATGA ACACAAGCAT TATCACCATC 29340  
 20 GTCGCTTGAT AAATAAGTAC TAGCAACTTT tTCGGAAATA TCTCAATAAG GTAACCAGAT 29400  
 ATAATGGACA ATGGAAATTT nAGAACCACT AAACCAACAA GATATATACC GACAATTGAT 29460  
 25 TGACTTAACA TATCTGTTAA ATATAGTGCT ATAAACGGTA TAAATGCTGT CGTAATAATT 29520  
 AGCTGTAAAA nATTGCTAAT CAATCGTACT TTCAA 29555

## (2) INFORMATION FOR SEQ ID NO: 207:

- 30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1539 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

AAAAAAAAAA AAAAAnGGTG AATCTTTAAT TAAACACTAA TATTGTAAAA GATGTTAAGT 60  
 40 AAACGCTTAA TGACACTTAT TTTTGTAAAA TAATAGTAAT ATCATTTTGT TAAATGAAAG 120  
 AATAAAGCTA TAATmATTAT AGAATAACTA TTAAAGGAG ATTATAAACA TGCCAATTAT 180  
 45 TACAGATGTT TACGCTCGCG AAGTCTTAGA CTCTCGTGGT AACCCAACTG TTGAAGTAGA 240  
 AGTATTAAGT GAAAGTGGCG CATTGCTGCG TGCATTAGTA CCATCAGGTG CTTCAACTGG 300  
 TGAACACGAA GCTGTTGAAT TACGTGATGG AGACAAATCA CGTTATTTAG GTAAAGGTGT 360  
 50 TACTAAAGCA GTTGAAACG TTAATGAAAT CATCGCACCA GAAATTATTG AAGGTGAATT 420  
 TTCAGTATTA GATCAAGTAT CTATTGATAA AATGATGATC GCATTAGACG GTACTCCAAA 480

55

AGCTGACTTA TTAGGTCAAC CACTTTACAA ATATTTAGGT GGATTTAATG GTAAGCAGTT 600  
 ACCAGTACCA ATGATGAACA TCGTTAATGG TGGTTCTCAC TCAGATGCTC CAATTGCATT 660  
 5 CCAAGAATTC ATGATTTTAC CTGTAGGTGC TACAACGTTT AAAGAATCAT TACGTTGGGG 720  
 TACTGAAATT TTCCACAAC TAAAATCAAT TTTAAGCAAA CGTGGTTTAG AAAGTGCAGT 780  
 AGGTGACGAA GGTGGTTTCG CTCCTAAATT TGAAGGTACT GAAGATGCTG TTGAAACAAT 840  
 10 TATCCAAGCA ATCGAAGCAG CTGGTTACAA ACCAGGTGAA GAAGTATTCT TAGGATTTGA 900  
 CTGTGCATCA TCAGAATTCT ATGAAAATGG TGTATATGAC TACAGTAAGT TCGAAGGCGA 960  
 ACACGGTGCA AAACGTACAG CTGCAGAACA AGTTGACTAC TTAGAACAAT TAGTAGACAA 1020  
 15 ATATCTATC ATTACAATTG AAGACGGTAT GGACGAAAAC GACTGGGATG GTTGGAACA 1080  
 ACTTACAGAA CGTATCGGTG ACCGTGTACA ATTAGTAGGT GACGATTTAT TCGTAACAAA 1140  
 CACTGAAATT TTAGCAAAAG GTATTGAAAA CGGAATTGGT AACTCAATCT TAATTAAAGT 1200  
 20 TAACCAAATC GGTACATTAA CTGAACATT TGATGCAATC GAAATGGCTC AAAAAGCTGG 1260  
 TTACACAGCA GTAGTTTCTC ACCGTTTCAGg aAACAGAAGA TACAACAATT GCTGATATTG 1320  
 25 CTGTTGCTAC AAACGCTGGT cAAATTAAAA CTGGTTTATT ATCACGTACT GACCgTATTG 1380  
 CTAAATACAA TCAATTATTA CGTATCGAgA TGAATTATTT GAAACTGCTA AATATGACGG 1440  
 TATCAAATCA TTCTATAACT TAGATAAATA ATTTCTnTA TAATCAAATG CTGACATAAT 1500  
 30 TTTAGTTGAG GATTATTATG ACGGTATAAA TAAATAAAG 1539

## (2) INFORMATION FOR SEQ ID NO: 208:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 846 base pairs  
 35 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

CAATTTCTAT CTATCAATGA TGTGCATACT TCCAnTTAAA TTAAtCGAAA TGaATCAAGG 60  
 TATATCATTCT CTGCCTCTTT ATATAACaAC AAATAGTGAT TACAATATTT CGGTTATTAA 120  
 45 CACGAAAATT TTACAAGCAC CTATTTTATT TACATATATA TACAGCAAAA AAGAAAGCCC 180  
 AGAAATATTG GTGTTTATTA AATCATTAA AAAGTATATT GCCAATGAAC AATTATAATA 240  
 AATTTCAAAT CTAAAAAACC AAGAATGCGA TTAATCATCA CATTCTTGGT TCAATTTTAT 300  
 50 TCATGAATTT TTTCAACATT AAACGTTAAG TTATTGTCTG AATTAAATT AACTTTAATC 360

55

CGTTGTACAA AACGTTTAA TGGTCTTGCA CCGTATTGAG GTTCATAAGC TTCTTGACCT 480  
 AGCCAAGCTT TAGCATCATC AGAAACTTCA ATTGAGATTC GTTGTTCTAA TAATCTTATA 540  
 5 TTTAATTGCG TTAAGATTTT ATCTACAATC ATACTCATGT CATCAATAGA TAATGGTTTA 600  
 AATAATACGA TATCATCCAT ACGATTCAAA ATTTCTGGTT TGAAATATGC ATTTAAACTT 660  
 GTCATAACAG CTTTTTCTGT TGATTCTGTA ATTTCAACCAG TCTCTTTTAC GTTTTCTAAT 720  
 10 AAAACTTGAG ATCCAATATT ACTTGTGATA ATAATAATAG TATTTTTTAA ATCAACGCTA 780  
 CGTCCTTTAG AATCAGTTAA ACGGCTTCAT CTAAAATTTG CAATAATACA TTAAAGACGT 840  
 CAGTAT 846

15 (2) INFORMATION FOR SEQ ID NO: 209:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1674 base pairs  
 (B) TYPE: nucleic acid  
 20 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

nTGGAACAG TAAGCCAGTA TTTTtagAAA GTTGCCATAC ATGAGCGTCG ATTTTCCAA 60  
 TATGGCTATG AACTAGAAC AATGGGAATT TGGAGGAAAA GTAAATGATT AAACCTAAAA 120  
 30 TAGCATTAAc CATTGCAGGT ACTGATcCaA CAGGTGGTGC CGGCGTAATG GCTGATTTAA 180  
 AATCATTTCA TTCATGTGGT GTATATGGTA TGGGCGTCGT TACAAGTmTT GTTGCTCAAA 240  
 ATACATTGGG CGTACAACAT ATTCATAATT TAAATCATCA ATGGGTAGAT GAACAACTTG 300  
 35 ATAGTGCTCT CAATGATACC TTACCTCATG CTATTAAAAC GGGGATGATT GCTACAGCAG 360  
 ATACTATGGA AACGATTTCGT CATTATTTAA TGCAACATGA ATCTATTCCA TATGTAATtG 420  
 ATCCTGTTAT GTTGGCGAAA rCggTGATTc ATaATGGwTA ATGACaCAAg CaAAACTTGC 480  
 40 AGCAtaCGTT ATTGCCATTA GCTGACGTAG TAACACCGAA TTTACCAGAA GCTGAAGAAA 540  
 TAACGGGACT AACCATTGAT AGTGAAGAAA AAATTATGCA GGCTGGCCGC ATCTTTATTA 600  
 ATGAGATTGG TAGTAAAGGT GTCATCATTA AAGGCGGTCA TTCAAATGAT ACTGATATAG 660  
 45 CAAAAGATTA TTTATTTACT AACGAAGGTG TTCAAACATT TGAAAATGAA CGATTTAAAA 720  
 CAAnACATAC GCATGGAACA GGGTGTACAT TTTCAgCAGT TATAACGGCA GAACCTGCAA 780  
 50 AAGGTAGACC ATTATTTGAG GCTGTACACA AGGCTAAAAA GTTTATTTCA ATGAGTATAC 840  
 AATATACGCC TGAAATCGGC CGTGGTAGAG GTCCAGTGAA TCATTTTGCA TATTTAAAGA 900

55

TGTATACAA ACGATGTAGT TAAAAATTTT ACAGCGAATG GTTTATTAAG TATTGGTGCT 1020  
 AGCCCTGCAA TGAGTGAAGC TCCCGAAGAA GCTGAAGAAT TTTACAAAGT TGCACAAGCG 1080  
 5 CTATTAATCA ATATCGGTAC TTTAACAGCA GAAAATGAAC AAGATATTAT TGCGATTGCT 1140  
 CAAACGGCAA ATGAGGCAGG CTTACCTATT GTATTTGACC CTGTAGCTGT TGGTGCTTCT 1200  
 ACATATCGAA AGCAATTTTG TAAATTATTA TTGAAATCAG CGAAAGTATC AGTAATTAAA 1260  
 10 GGCAATGCAT CTGAAATATT AGCGTTGATT GATGATACAG CAACTATGAA AGGTACAGAT 1320  
 AGTGATGCTA ATCTTGATGC GGTTGCAATA GCGAAAAAGG tTACGCAACA TATAAACTG 1380  
 15 CAATAGTAAT CACAGGTAAA GAGGACGTTA TTGtTcMAGA TAATAAAGCC TTCGTATTAG 1440  
 CTAATGGATC TCCATTATTA GCACGAGTAA CTGGAGCTGG TTGTTTATTA GGAGGCGTTA 1500  
 TTGCTGGATT TTTATTTAGA GAAACAGAAC CAGACATAGA AGCGTTAATT GAAGCGGTAA 1560  
 20 GCgkATTTAA TATTGCTGCT GAGGTAGCTG CTGAAAATGA AAATTGTGGT GGTCTGGTA 1620  
 CGTTTTACC ATTGTTGCTT GATACGTTAT ATCATTTAAA TGAAACAACC TATC 1674

(2) INFORMATION FOR SEQ ID NO: 210:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2232 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

35 ATGAGTTGCC GATGAATTTA GCACCACCAA CGATTGChTT TGATACTGTG TCCCAACCAG 60  
 CTGTGTTAGC ATATTTAATA CCTTCACGTA AAGGATCGTT ATCATATGCA GCAATACCAA 120  
 ATACGTTATG GTATTTCTGT TTTGAGTTAG TTACAACTTT GTnTTGCACT ACATCTGCAC 180  
 40 CTTTcGCTAA TTGAGAAGTA CCGTTACCTG TTTCTAATAG GGCATGTGAG ATAAGATAAA 240  
 CTTCAATTAAT GCCATACATT TGAGCAGCTT TGTTAAATGC AGCACCTTGG TTTTCTAATA 300  
 CACCTTTACC TTTTAAGAAT TGATTAATTT TATCAATAGA AATATTTTGT GGTGTTGCTA 360  
 45 AGCGTAAGAA TTGATATTTT AATGCTGGAT CTTGAGCTAA ACGCTTCGTA TCCATTGCAT 420  
 GCTTAACATC ATTAAATTTA GCATCTGTCC ACTTACCTGG TACACGTTGT ACTTGTGGTT 480  
 50 TATATTGTAA ACCAGCTTGT ATTTGAGCAA CTTGGTTTAA TGTCATACCT GTTTGATTAT 540  
 ACTTAATTAA TTCTTTAGCT AAATCAGTTG ATTTAATCCA TGCTAAITTA CCGTTAGATA 600  
 ATTTACCATA GTACCAAGTT TGTCCATTAA TGACTTGTTT TTTAACAACG GCGAATGGTT 660  
 55

AACCATTACC ATTTTAAATT ACATAAGTGT AGTTATAATC TTTGGCAGCT GATGTAGTTG 780  
 GTTTCACAGC AGTTGGTGCA GTTAAATCTT TTGCATTTAC CCAACCAGTG CGGTTATTAA 840  
 5 TAGTACCGTA TAAATAAACA TCTTTGCCTA CAGATACTTG TTTCGTTGCA TTAAATGTAC 900  
 CTTGAGCAAT GTTATTGCCT GTTAAATGA CTTGGTTTTT AGTACCCCAA GGAACCATTG 960  
 10 ATAAGCCGTT ATTTGATTTA TTAACAGTAT ATTTTGTAGT CGTTTTAACT TCTTTGCCTA 1020  
 AGTTTTGAAC ATTTAAGTCT TTTACATTGA ACCAACCTAA TGGGATGTTA TGGCTTGTAT 1080  
 TGTTTAATAA TACATACGTT TCATTACCAT GAGCACGCTC TTTTGTTACA TAGAACGTAC 1140  
 15 GGTCTGCATA TTTCGCACCG TTTTTCGCTG TTTTTCATA AACAGAAGCA CGAATACCAG 1200  
 TGTGTGTTGG TTTAACTTGA GCAATCTTGC TAACTGTTTG AGTCGTTTGT GGTTTAGTAA 1260  
 CAGTATAAGC TTTTACAGCT GTTTTTGGTT GTGCTACTGC TTTTTTAGGT GCAGCAGGTA 1320  
 20 CAGCTAAATA TGCTTTACTT ACCCAACCAG ATTTACCATT TACAGTTCCA AATAAATAGA 1380  
 TAGATTTATC AATTTGTTGT TGCTTAGTCG CTTTAAAAGT TTGGTTACCT GTACCAGAAA 1440  
 25 CTGCACCAGC TTCTTGTTTA TAAGTGCCCC AAGGTACTGA ATATAATTTA GTGCCTGGgT 1500  
 TTAGTGATA TGTTTGCAAT ACATTTACAG GTGATTTTGC ATtGtTATAA ATACGTCACC 1560  
 TTGTTTAACC CAACCAATTA AAGTTGGACT ATTGTAATCT TTAACATAAGT AGAATTTGTT 1620  
 30 TCCACCTAAA CTTGCTTCTT TTGTTACAGC AAATGTTTTT TGAACCTCTT TCGTTGGCTT 1680  
 ACCAGTTTGT TCATAAACTG TAGTGAATAA GCCATTGTTT TTAGCATTAA TTTGAGCAAC 1740  
 ACCGTTTAAT GATGAAACTG TTAATTTATT ATTTGTTGTA GGTGTTGATG GCTTAGGTGT 1800  
 35 TGGTGTAGGC GTAGGTTTAG CAGTATCAAC TAAATATGCT TTAATTACCC AACCAGATTT 1860  
 ACCATTCACA GAGCCATATA AATAAATTGA TTTATCAATT TGTGTTGCT TTGAAGCCTT 1920  
 40 AAATGTTTGG TTTCCAGAGC CAGACACACT ACCAGCAACT TGTTTAGATG TACCCCAAGG 1980  
 TACTGTATAA AGTTTCGTAC CAGGTTTGAT TGAATATGAT TGATTTACAT TTACAGGTGA 2040  
 TTTAGCTGTG TTGTAAACCA CATCGCCTTC TTTAACCCAA CCAAATTTAT TACCAGAATT 2100  
 45 GTAATCTTGA ACAAGATAGA ATTTTGTATT ACCTAATGTA GCTGTTTTAG ATACAGCAAA 2160  
 TGTTTTTTGA ACTTCATTAG TTGCTTTACC AGTTTTGTCTG TATACAGTAG TATATAAACC 2220  
 ACTATTTGTT GG 2232

50

(2) INFORMATION FOR SEQ ID NO: 211:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2082 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

5	GATTTAAATA AAATTAATGG ATATCGTGAT CGTACGATGT TAGAACTTCT GTACGCAACG	60
	GGAATGCGTG TATCTGAATT GATACATTTA GAGTTAGAAA ACGTGAACCTT AATAATGGGA	120
	TTTGTACGCG TATTTGGTAA AGGCGATAAA GAAAGAATTG TACCATTAGG CGACGCAGTC	180
10	ATTGAGTACT TAACTACTTA TATTGAAACG ATTAGACCGC AACTTTTAAA AAAGACTGTT	240
	ACTGAAGTCT TATTTTAAA TATGCATGGT AAACCTTTAT CACGACAAGC AATATGGAAA	300
	ATGATTAAAC AAAATGGTGT AAAGGCAAAC ATTAAAAAGA CGTTAACGCC ACATACGTTA	360
15	CGCCACTCTT TTGCGACACA TTTATTGGAA AATGGCGCAG ATTTAAGAGC AGTGCAAGAG	420
	ATGTtAGGtC ACTCTGaCmT ATCTACTACC CmaCTCTATA CmCATGTTTC GrAATCTCAA	480
	ATTAGAAAAA TGTATAACCA ATTTATCCTT AGAGCATAAA GTGAACAATA ACTCAAAAGT	540
20	CACAATACAC ATGACTAAAA ATGTCTGTGC TATTGTGGCT TTTTAAATT GGTGATTAA	600
	TTACGTCTAT GTTTTCTTAA TTGAATCGCT TCTTCTTTTG CTGCAATCAC TTCTGAACGA	660
25	TCACGGCGCA TGTGATGGTC TACAATAAAA GGATCTGTTG CTGTTTCCTG ATTATAATCA	720
	TAGTCTGGAT AGTTGGCCTT GATGATGCGT TCAAAGACTG GAGTTATTGG TAATATAACA	780
	GATGAAAAAG GCTTTGCTGC ATTCAATTTT GCAATCTGTT GCTCAATTAA CAACTGATAA	840
30	TCATTTAAAT TAAGGTATAA CGCATCTCTA TCTTTAGCAT TTTGTATTAT TTCTTTAGAT	900
	TTATTTAAAG ACTTATAGGC GCCTTTTAAA TTATTGCGGC GATAATGGTA ACAAGCAGTT	960
	GCAAACAAGA TTAAACTAAC AACTGCATCT TGCTTACTGT AGTTATTTTC AGCTTTCCAT	1020
35	GCATCTTCTA AAATGTCATG ACATAGGAAA TAATGTTGCT TAGTATGAAA TTGATAATAG	1080
	AAAF <sup>~</sup> TTATCA GTGCCTGTTG CATTTTGTTA TCACCCCAAT TTAAAAGTAA GTTATTTTCA	1140
40	TGCTATAATA TTTTAGAGAA TTATGCACAT ATGACGCAAT ACGAGGTAGA TATTATGTAT	1200
	GAAGTTAAAT TAGATGCTTT CAATGGACCA TTAGATTTAT TGCTGCATCT TATCCAAAAA	1260
	TTTGAAATAG ATATTTATGA TATTCCTATG CAAGCATTAA CAGAGCAGTA TATGCAGTAC	1320
45	GTTTCATGCAA TGAAACAGCT TGAAATTAAT ATTGCAAGTG AATACCTAGT ATTAGCGTCA	1380
	GAAC <sup>~</sup> CTTAA TGATTAAAAG TAAGATGCTA TTACCACAAT CAACATCAGA TATGGATGTT	1440
	GATGATGACC CACGGGAAGA TTTAGTtGGG CGTTTAAATAG rATATCaAAA TTATArAGAA	1500
50	TATACTGctA TTTTAAATGA CATGAAAGAA GAAAGAGATT TTTATTTTAC CAAAAAGACC	1560
	GACAGATTTA TctCATTTGG AAaCAGATGA ATCyTGGGAT CCaAATCATA CGATTGATTT	1620

55



ATCTGTTGAA ATCCGAAAAG AGACATTAC CATTCAACAA GCTACAGAAC AAGTGACATC 1740  
 GAGATTGAAA GATAAAGATC ATTTTAACTT CTTTAGTCTG TTTACGTTTT CTGAGCCAAT 1800  
 5 TGAACAAGTA GTCACTCACT TTTTAGCTAT TTTAGAGATG TCAAAAGCAG GAATAATTAA 1860  
 TATTGAGCAA CAACGTAATT TTGAAGATAT TAACATTATT AGAGGAGTGA ACTACCATTT 1920  
 10 TGGATAATCA TGGTATATTA GAGTCGCTTT TATTTACAGC TGGCGATGAA GGTTTAGATG 1980  
 AAAACAACCT ATTAGAAATA TTAGATATGT CGAAAGACCA ACTCGTTGAA TTAATTGAAA 2040  
 ATTATTCATC ACATGGATTA ATGATACAAC GATTTGGAAT GA 2082

15 (2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4219 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

25 TCTATTCTCG TTCTTCCAAG ACCCTGaATT AGAAGTTAAG AAAATCGAAG AAGATGAGAA 60  
 AGAATCTATT AAAAAAGCTC AAAAAGGTAT TTATAAAGAC CCTAGAGACA TCAATGATGA 120  
 30 CGAACAAGAT GATGATACAA AAGATACTGT TGATAAAAAG GAATGATTGT AATTGCCTAA 180  
 CAAAAACACT CAAGAATATT GGGAAGAACG CGGACGCAAA GCAATCGAGA ATGAGTTGAA 240  
 GCGTGATAAA ACTAAAGCTG AAGAAATAGA ACGTATATTG AATATGATGA TTAAGCGCAT 300  
 35 TGAAAAAGAG ATCaATGCGT TTATTGTCAA GTACGGAGAT TTTGCAGGCG TTACATTACA 360  
 AGAAGCACAA AAGATTATTG ATGAGTTCTGA TGTAAAAGCG TTTCAAGAAG AAGCAAAAAG 420  
 ATTGGTCGAA AACAAGGAGT TTAGCGATAG AGCAAATGAA GAATTAAAGA AGTATAACAC 480  
 40 GAAAATGTAT GTATCTAGAG AACAGATGTT AAAGATTCAA ATAGAATTCT TAATTGCTTA 540  
 TGCAACAGCT CAAACAGAAT TATCGATGAG GGAATATTTT GAATCAACAG CTTATCGTGT 600  
 GTTCAGTGAT CAAGCGGGTA TTTTAGGTGA AGGTGTACAA GTAGCTAAAG AAGTTATAGA 660  
 45 TACAATCGTT GATACACAAT TTCATGGTGT CGTTTGGTCA GAGCGATTAT GGACTAATAC 720  
 CGAAGCAATG AAACAAGAAG TAGAAGAAAT AATTGCTAAT GTAGTTATTA GAGGTCGACA 780  
 50 TCCTAATGAA TATGTTAAAG ATATGCGCAA CACTTAAATA AATTCGAAGG CACAGCACGA 840  
 CAAAAGACCG CAGCAATTAA ATCATTGCTT TATACGGAAT CGGCACGTGT TCACGCACAA 900  
 TCAAGCATTG ACAGCATGAA AGAAATTTCA CCGGAAGgAT ATTATATGTA TATTGCAAAA 960

55

	GACGCTAAAA TTGGTGTTAA TTTCTATCCT ATGCATATCA ATTGTCGTTT AGATTGCGCT	1080
	TTACTACCTA AATCTATGTG GCCGAAAAAA CCAAGCAAGA AACGAAAAAC AAAATACTTC	1140
5	GGAGGGAAAG TGAAAAGCGG TGATTGATTT AAAAGTGAAG TTTTTTAAAG GCAAGTTAGT	1200
	TTTGTATGAC AGTAAATTAA ATGTTTGGAG GATACTAATA TGAGTAATAC TGACAAATAC	1260
	CTTAGAGACA TAGCAAGAGA ATTAAAAGGT ATACGTAAAG AGTTACAAAA GCGAAACGAA	1320
10	ACAGTTATTA TTGATGCAAA CTTAGACAGT TTAAGGTCGG CAGTATTAGC CGATAAAGAA	1380
	AAATCGAAAT ATAATGAACC TCTCTTTTAA TAGCTAGCAC TTAATTGTGT TGGCTATTTT	1440
15	TTATGTCCAA AACGTGCTGA TGACATAAAA AGCACGCATG GAAAAACAGT CGACAGACTA	1500
	TAAATGGAGG TATATCTCAT GGAAGAAAAT AAACCTAAGT TTAATTTGCA aTTTTTTGCA	1560
	GACCAATCAG ATGATCCGGA CGAACCAGGC GGAGATGGTA AAAAAGGAAA TCCTGATAAG	1620
20	AAAGAAAATG ACGAAGGTAC TGAAATAACT TTCACGCCAG AGCAACAAAA GAAAGTTGAT	1680
	GAAATACTTG AACGTCGTGT AGCCCACGAA AAGAAAAAAG CTGATGAGTA TGCAAAAGAA	1740
	AAAGCAGCAG AAGCTGCTAA AGAAGCTGCT AAATTAGCGA AAATGAACAA GGATCAAAAA	1800
25	GATGAATATG AACGCGAACA AATGGAAAAA GAACTGGAAC AATTACGTTT AGAAAAACAA	1860
	TTAAACGAAA TCGGTTTCTA AGCACGAAAA ATGTTGAGTG AAGCGGAGT TGATTCATCA	1920
30	GATGrGGTTG TCAATTTAGT TGTAACAGAT ACTGCTGAAC AAATAAATT GAATGTTGAA	1980
	GCTTTTTCTA ATGCAGTAAA AAAAGCGGTT AATGAAGCGG TTAAGGTTAA CGCTAGACAA	2040
	TCGCCATTGA CTGGTGGAGA TTCATTTAAT CACTCGACTA AAAATAAACC GCAAAACTTA	2100
35	GCTGAAATAG CTAGACAAAA AaGAATTATT AAAAATTAAC GGAGGCATTT AAATGGAACA	2160
	AACACAAAAA TTAAATTTAA ATTTGCAACA TTTTGCAAGT AACAATGTTA AACCACAAGT	2220
	ATTTAACCCT GACAATGTAA TGATGCATGA AAAGAAAGAT GGCACGTTGT TAAACGACTT	2280
40	TACAACACCT ATCTTACAAG AGGTTATGGA AAACCTCTAA ATCATGCAAT TAGGTAAGTA	2340
	CGAACCAATG GAAGGTACTG AGAAGAAGTT TACTTTTTGG GCTGATAAAC CAGGTGCTTA	2400
	CTGGGTAGGT GAAGGTCAAA AAATCGAAAC GTCTAAGGCT ACTTGGGTTA ATGCTACAAT	2460
45	GAGAGCGTTT AAATTAGGGG TTATCTTACC AGTAACAAAA GAATTCTTGA ATTACACTTA	2520
	TTCACAATTC TTTGAAGAAA TGAAACCTAT GATTGCTGAA GCTTTCTATA AAAAGTTTGA	2580
50	CGAGGCAGGT ATTTTGAATC AAGGTAACAA TCCGTTCCGT AAATCAATTG CACAATCAAT	2640
	TGAAAAAACT AATAAGGTTA TTAAAGGTGA CTTACACAA GATAACATTA TTGATTTAGA	2700
55	GGCATTGCTT GAAGATGACG AATTAGAAGC AAATGCATTT ATCTCAAAAA CACAAAACAG	2760

TGATTCGTTA GACGGTCTAC CTGTGGTTAA CCTTAAATCA AGCAACTTAA AACGTGGTGA 2880  
 ATTAATCACT GGTGACTTCG ACAAATTGAT TTATGGTATC CCTCAATTAA TCGAATACAA 2940  
 5 AATCGATGAA ACTGCACAAT TATCTACAGT TAAAAACGAA GATGGCACAC CTGTAAACTT 3000  
 GTTTGAACAA GACATGGTGG CATTACGTGC AACTATGCAT GTAGCATTGC ATATTGCTGA 3060  
 TGATAAAGCG TTTGCTAAGT TAGTTCCTGC TGACAAAAGA ACAGATTTCAG TTCCAGGAGA 3120  
 10 AGTTTAATAA ATAATTAGGA GTGGTAACAT GCCCGAAATC ATTGGAATTG TTAAAGTAGA 3180  
 TTTTACAGAT TTAGAAGATA ACAGACATGT CTATATGAAA GGGCATGTCT ACCCTCGTAA 3240  
 15 AGGTTATAAT CCTACAGATG AACGTATCAA AGCTTTAGCT AGTGTTGAAA ATAAACGCAA 3300  
 CAAACAAATG ATTTACATTG TAAATGACAA ATTAACCAAA AAAGAACTTG TCGAAATAGC 3360  
 AAGTGTGCTT GGCTTACAAG TTGATGAAAA ACAACAAAA GCTGAAATTA TCAATGCTTT 3420  
 20 TGAGTCACTA GAGTAGGTGG TTATATGACT ACGCTAGCTG ATGTAAAAAA ACGTATTGGT 3480  
 CTTAAAGATG AAAAGCAAGA TGAACAATTA GAAGAAATCA TAAAAAGTTG TGAAAGCCAG 3540  
 TTGTTATCAA TGTTACCTAT TGAAGTTGAA CAAATACCGG AAAGgTTTAG TTACATGATT 3600  
 25 AAAGAAGTTG CAGTTAAACG CTACAACAGG ATTGGTGCTG AAGtATGACA TCAGAAGCGG 3660  
 TTGACGGACG TAGCAATGCG TATGAATTGA ACGATTtCAA GGAGTATGAA GCTATTATTG 3720  
 ATAATTACTT TAATGCTAGA ACGAGAACTA AAAAAGGAAG GGCTGTGTTC TTTTGAGATA 3780  
 30 TGAAGATAGA GTTATTTTTC AATTAGAACA AGTAGCAACT TACAATCCTA AAAC TAGCAA 3840  
 AAAAGAAAAC AACTAATCA CTTATGATGC GATACCATGC AATATTAACC CCATTTCTAG 3900  
 35 AGCAAGAAAG CAACTTGAAT TTGGTGATGT AAAAAACGAT GTAAGTGTTT TGAGGATAAA 3960  
 AGAATCAATA TCTTACCCTG TTAGCCACGT GTTGGTTAAT GGCATTTCGCT ACAAGATAGT 4020  
 TGATāCAAGG ATATACAGAC ACGAAACGTC ATATTATATC GAAGAGGTCA ATTGATGAAT 4080  
 40 ATAGATGGAT TAGACGCACT GTTAAACCAA TTTCACGATA TGAAAACCAA CATTGATGAT 4140  
 GATGTAGATG ATATTTTACA GGAAAACGCC AAAGAATATG TAGTACGAGC TAAATTGAAA 4200  
 GCTAGAGAAG TAATGAATA 4219

(2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1999 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

	GCTTACAAGT ATATTCATAA TTACATATTC AAGGTCCTTG CATGTGGTAT TTTGCTATGG	60
	yCtTTaACTA CAACGGGGTC TAAGACTGCG TTTATCATAT TAATCGTCTT AGCCATTtAT	120
5	TyCTTTATka AAAAGTTATT TAGTAGAAAT GCGGTAAGTG TTGTGAGTAT GTCAGTGATT	180
	ATGCTGATAT TACTTTGTTT TACCTTTTAT AATATCAACT ACTATTTATT CCAATTAAGC	240
	GACCTTGATG CCTTACCGTC ATTAGATCGA ATGGCGTCTA TTTTGAAGA GGGCTTTGCA	300
10	TCATTAAATG ATAGTGGGTC TGAGCGAAGT GTTGTATGGA TAAATGCCAT TTCAGTAATT	360
	AAATATACAC TAGGTTTTGG TGTCGGATTA GTGGATTATG TACATATTGG CTCGCAAATT	420
15	AATGGTATTT TACTTGTTGC CCATAATACA TATTTGCAGA TCTTTGCGGA ATGGGGCATT	480
	TTATTCGGTG CATTATTTAT CATATTTATG CTTTATTTAC TGTTTGAATT ATTTAGATTT	540
	AACATTTCTG GGAAAAATGT AACAGCAATT GTTGTAAATGT TGACGATGCT GATTTACTTT	600
20	TTAACAGTAT CATTTAATAA CTCAAGATAT GTCGCTTTTA TTTTAGGAAT TATCGTCTTT	660
	ATTGTTCAAT ATGAAAAGAT GGAAAGGGAT CGTAATGAAG AGTGATTCAC TAAAAGAAAA	720
	TATTATTTAT CAAGGGCTAT ACCAATTGAT TAGAACGATG ACACCACTGA TTACAATACC	780
25	CATTATTTCA CGTGCATTTG GTCCAGTGG TGTGGGTATT GTTTCATTTT CTTTCAATAT	840
	CGTGCAATAC TTTTGTATGA TTGCAAGTGT TGGCGTTCAG TTATATTTTA ATAGAGTTAT	900
30	CGCGAAGTCC GTTAACGACA AACGGCAATT GTCACAGCAG TTTTGGGATA TCTTTGTCAG	960
	TAAATTATTT TTAGCGTTAA CAGTTTTTGC GATGTATATG GTCGTAATTA CTATATTTAT	1020
	TGATGATTAC TATCTTATTT TCCTACTACA AGGAATCTAT ATTATAGGTG CAGCACTCGA	1080
35	TATTTTCATGG TTTTATGCTG GAACTGAAAA GTTTAAATTT CCTAGCCTCA GTAATATTGT	1140
	TGCGTCTGGT ATTGTATTAA GTGTAGTTGT TATTTTTGTC AAAGATCAAT CAGATTTATC	1200
	ATTGTATGTA TTTACTATTG CTATTGTGAC GGTATTAAAC CAATTACCTT TGTTTATCTA	1260
40	TTTAAAACGA TACATTAGCT TTGTTTCGGT TAATTGGATA CACGTCTGGC AATTGTTTCG	1320
	TTCGTCATTt AGCATACTTA TTACCAAATG GACAGCTCAA CTTATATACT AGTATTTCTT	1380
45	GCGTGTGTTCT TGGTTTAGTA GGTACATACC AACAAAGTTGG TATCTTTTCT AACGCATTTA	1440
	ATATTTTAAC GGTCGCAATC ATAATGATTA ATACATTTGA TCTTGTAATG ATTCCGCGTA	1500
	TTACCAAAT GTCTATCCAG CAATCACATA GTTTAACTAA AACGTTAGCT AATAATATGA	1560
50	ATATTCAATT GATATTaCA ATACCTATGG TCTTTgGTTT AATTGCaATT ATGCCATCAT	1620
	TTTATTTATG GTTctTTGGT GAGGAATTCG CATCAACTGT CCCATTGATG ACCATTTTAG	1680
55	CGATACTTGT ATTAATCATT CCTTTAAATA tGTTGaTAAG CaGGCAATAT TTAtTAAtAG	1740

TATGTATAT TTTGATATAT TTTTATGGAA TTTACGGTGC TGCTATTGCG CGTTTAATTA 1860  
 CAGAGTTTTT CTTGCTCATT TGGCGATTTA TTGATATTAC TAAATCAAT GTGAAGTTGA 1920  
 5 ATATTGTAAG TACGATTCAA TGTGTCATTG CTGCTGTTAT GATGTTTATT GTGCTTGGTG 1980  
 TGGTCAATCA TTATTTGCC 1999

(2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7769 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

20 TCATTATTAA GACTATTATA TATAATGAAT TTAACTGGT TTATTAAACG AGAACGTCGG 60  
 GAATTAAGTA ACTACAATAA AAATAAGATA TGACAATAAG GAGACTACAC GCGTGATCAT 120  
 TGCCATAATT ATATTGATAT TTATTTTCGTT TTTCTTTTCA GGAAGCGAGA CGGCATTAAAC 180  
 25 GGCTGCCAAT AAAACAAAT TTAAACTGA AGCTGACAAA GGTGATAAAA AAGCAAAAGG 240  
 CATTGTAAAG TTACTTGAAA AACCAAGTGA GTTTATTACA ACGATTCTAA TTGGGAATAA 300  
 TGTCGCGAAT ATTTTATTAC CAACACTTGT TACAATTATG GCTTTACGTT GGGGGATTAG 360  
 30 CGTTGGTATT GCATCAGCTG TTTTAACAGT TGTTATCATT TTGATCTCCG AAGTGATTCC 420  
 CAAGTCTGTC GCTGCAACAT TTCCAGATAA AATAACAAGG CTTGTATATC CAATTATTAA 480  
 TATTTGTGTC ATTGTGTTCC GTCCTATCAC ATTACTTTTA AATAAGTTGA CGGACAGTAT 540  
 35 TAATCGAAGT TTATCTAAGG GCCAACCTCA AGAACATCAA TTTTCAAAG AAGAATTTAA 600  
 AACAAATGTTA GCAATTGCTG GACATGAAGG TGCTTTAAAT GAAATTGAGA CGAGTAGGTT 660  
 40 GGAAGGTGTC ATTAATTTTG AAAATTTTAA AGTAAAAGAT GTAGATACAA CACCTAGAAT 720  
 TAATGTGACG GCATTTGCTT CAAATGCGaC ATACGAAGAA GTTTATGAAA CGGTTATGAA 780  
 TAAGCCATAC ACTAGATATC CAGTGACGA GGGAGATATT GATAACATTA TTGGGGTGTT 840  
 45 TCATTCTAAA TATCTGTTGG CTTGGAGTAA TAAAAAGAA AATCAAATTA CAACTATTC 900  
 AGCTAAGCCA TTATTTGTGA ATGAACACAA TAAAGCTGAA TGGGTATTAC GTAAGATGAC 960  
 TATTTCTAGA AAACATTTAG CAATTGTGTT GGACGAATTT GGTGGTACTG AAGCGATAGT 1020  
 50 GTCACATGAA GACTTAATTG AAGAATTATT AGGTATGGAA ATTGAAGATG AGATGGATAA 1080  
 AAAGGAAAAA GAAAACTTT CTCAACAGCA AATTCAATTT CAACAACGGA AAAATCGCAA 1140

	GTATTGAATA TCCAATTATA CAAGCAGGTA TGGCAGGAAG TACGACACCG AAATTAGTTG	1260
	CATCAGTAAG TAACAGTGGT GGGTTAGGCA CAATAGGCGC AGGTTACTTT AATACGCAGC	1320
5	AATTGGAAGA TGAAATAGAT TATGTACGCC AATTAACGTC AAATTCCTTT GGCGTAAATG	1380
	TCTTTGTACC AAGTCAACAA TCATATACCA GTAGTCAAAT TGAAAATATG AATGCATGGT	1440
	TAAAACCTTA TCGACGCGCA TTACATTTAG AAGAGCCGGT TGTAAAAATT ACCGAAGAAC	1500
10	AACAATTTAA GTGTCATATT GATACGATAA TTAAAAAGCA AGTGCCTGTA TGTGTGTTTTA	1560
	CTTTTGGAAT TCCAAGCGAA CAGATTATAA GCAGGTTGAA AGCAGCGAAT GTCAAACTTA	1620
15	TAGGTACAGC AACAGTGTT GATGAAGCTA TTGCGAATGA AAAAGCGGGT ATGGATGCTA	1680
	TCGTTGCTCA AGGTAGTGAA GCAGGTGGAC ATCGTGGTTC ATTTTAAAA CCTAAAAATC	1740
	AATTACCTAT GGTGGAACA ATATCTTTAG TGCCACAAAT TGTAGATGTC GTTCAATTC	1800
20	CGGTCATTGC CGCTGGTGGA ATTATGGATG GTAGAGGAGT TTTGGCAAGT ATTGTCTTAG	1860
	GTGCAGAAGG GGTACAAATG GGCACCGCAT TTTTAACATC ACAAGACAGT AATGCATCAG	1920
	AACTACTGCG AGATGCAATT ATAAATAGTA AAGAAACAGA TACAGTCATT ACAAAGCGT	1980
25	TTAGTGGAAG GCTTGACGCG GGTATCAACA ATAGGTTTAT CGAAGAAATG TCCCAATACG	2040
	AAGGCGATAT CCCAGATTAT CCAATACAAA ATGAGCTAAC AAGTAGCATA AGAAAAGCCG	2100
30	CAGCAAACAT CGGCGACAAA GAGTTAATAC ATATGTGGAG TGGACAAAGC CCGCGACTAG	2160
	CAACAACGCA TCCCGCCAAC ACCATCATGT CCAATATAAT CAATCAAATT AATCAAATCA	2220
	TGCAATATAA ATAATCGACC GCAATCCACA AAAGCACAAG CACCCCCAAA CATTATTTTA	2280
35	GTGCTTGCCA TTTTGTGGA TTGCGTTTCT ATTTTACCAA TTTAATCAAA CGAAAACATC	2340
	AAGCTGAAGA TCGCCGAAAG ATTTTAATCA AGCAAAAACA TCAAACTAAA GTTCGCTGAA	2400
	ATGATTATGA TAAAAGTTAT ATGGTATGAT GACATTGGTG ATATATATGA TAAACATCGG	2460
40	ATTAACAGGT TGGGGTGATC ACTATTCATT ATATGAAGAT TTAGAACGCC AAACCGATAA	2520
	ACTTAAAACA TATGCTGGAC ATTTTCCGGT TGTCGAATTA GATGCGACAT ACTATGCGAT	2580
45	ACAACCGGAA AGAAATATAT TGAAATGGAT AAAAGAAACG CCTGATACAT TTGAATTGT	2640
	GGTCAAAATT CATCAAGCAC TcACATTGCA TGCAGACTAC AAAACATTTG CAGATACAAG	2700
	GCAAGAACTA TTTGATCAAT TTAAGAAATAT GTTAGAGCCC TTACATACAC AGAAAAAATT	2760
50	AGCAATGGTA TTGGTTCAAT TTCCGCCATG GTTTGACTGC AATGCACAAA ATATCAAATA	2820
	TATTTTGTAT GTAAGACAGC AATTACAAGC ATTTCCAATG TGTGTAGAAT TTAGGCATCA	2880
	ATCATGGTTT AGTGATGCAT TTAAAGAACA AACATTGGCA TTTTAAACAG AACATCAAAT	2940
55		

	AATCACAAAT GAAATTGCGT TTGTACGTTA TCATGGACGT AATCATTACG GTTGGACTAA	3060
	GAAAGATATG TCAGATCAAG AATGGCGCGA TGTACGCTAT TTATATGATT ATAATGAGCA	3120
5	AGAATTAATA GACTTGGCAC AAAAGGCACA AATATTAGCA CAAAAAGCTA AGAAAGTTTA	3180
	CGTCATATTT AACATAATT CTGGTGGTCA TGCAGCAAAT AATGCCAAAA CATATCAGCG	3240
	ATTATTGAAT ATAGAATATG AAGGGTTAGC ACCACAACAA TTAAAATTAT TTAAAGAGGC	3300
10	GACGACTATG TTATTAACAA TTACATTATT AGTTTTAATC GGAGGTTTGT CAGCGATTAT	3360
	AGGGTCTATC GTAGGCATTG GAGGCGGTAT TATTATCGTT CCAACAATGG TTTACCTCGG	3420
15	TGTTGAACAT GGATTACTAC ATAATATTAC AACACAAGTA GCGATAGGGA CGTCTTCAGT	3480
	CATTCTAATT GTGACAGGAC TTTCTTCATC ACTTGGATAT TTAAAAACAA AACAAAGTTGA	3540
	TATTAAAAAT GGTTCCATCT TTTTATTTGG ACTATTACCA GGTTCAATTGC TTGGGTCCTT	3600
20	CATTAGTAGA TATTTAACAT TTGAGTCATT TAATTTATAT TTTGGTATCT TTTTAATTTT	3660
	CGTAGCCATT TTATTAATGG TAAGAAATAA GATTAAACCG TTTAAAATTT TCGATAAACC	3720
	CAAGTATGAA AAGACTTATG TAGACGCTAA AGGTAAAACA TATCATTATA GTGTTCCACC	3780
25	ATTGTTTGCT TTTATTACAA CGTTTTTAAT TGGTATATTG ACAGGTTTAT TTGGTATTGG	3840
	AGGTGGCGCA CTAATGACGC CACTAATGCT TATTGTATTT AGATTTCCAC CTCATGTAGC	3900
30	TGTTGGAACA AGTATGATGA TGATTTTCTT TTCAAGTGTC ATGAGTTCTA TAGGGCACAT	3960
	TGCTCAAGGT CACGTAGCTT GGGGTTATGC AATCATTTTA ATTATTTCTA GTTATTTTGG	4020
	TGCGAAAAATC GGTGTCAAAG TGAATCAATC AATTAAGTCA GATACGGTAG TAACATTATT	4080
35	GAGAACAGTA ATGTTGTAA TGGGTATATA TTTAATTATT CGTGCGTTGA TTTAATACAA	4140
	CTTTAAAAGG AGGACGTCAA TTTGAGGCTT ACAATTTATC ATACGAACGA TATTCATAGT	4200
	CATTACATG AATACGAACG CATTAAAGCA TATATGGCAG AACATCGGCC ACGACTTAAT	4260
40	CATCCTTCTT TATATGTTGA TCTAGGTGAT CATGTAGATT TATCCGCACC TATAACTGAA	4320
	GCAACTTTAG GTAAAAAGAA TGTGGCATT AATAATGAAG CAAAATGTGA TGTTGCAACA	4380
45	ATCGGTAATA ATGAAGGGAT GACCATTTC AACAAGCTT TAAATCACCT TTACGACGAA	4440
	GCAAAATTTA TAGTGACATG TAGCAATGTT ATAGATGAAT CAGGTCATTT ACCAAATAAT	4500
	ATCGTTTCTT CTTATATTAA GGACATAGAC GGTGTGAAAA TACTATTTCGT TGCAGCGACA	4560
50	GCACCTTTTA CCCCATTTTA TCGTGACTA AATTGGATTG TTACCGATCC ACTTGAATCT	4620
	ATAAAAGAAG AAATTGAACT TCAACGAGGT AAATTTGATG TATTAATCGT GCTAAGTCAT	4680
55	TGTGGCATT TCTTCGATGA AACATTATGC CAAGAATTGC CTGAAATTGA TGTCATTTTT	4740

	GCAGCTGGAA AGTATGGTAA TTATCTTGGG GAGGTTAATT TAACTTTTGA GGCACATAAA	4860
	GTAGTACATA AAAGTCAAAA GATTATTCCT TTAGAAACAT TACCTGAAGT TGAAACTTCA	4920
5	TTTGAAGAAG AAGGAAAAAC GTTAATGTCC AATTCAGTAA TTCAACATCC AGTAGTGCTT	4980
	AAGCGTAGTA TGAATCACAT AACTGAAGCT GCATACTTAT TAGCTCAAAG TGTTTGTGAG	5040
	TATACACATG CACAATGTGC CATCATCAAT GCTGGCTTAC TCGTTAAAGA TATTGTAAAA	5100
10	GATGAAGTGA CAGAATATGA CATTTCATCA ATGTTACCGC ATCCGATTAA TATGGTAAGG	5160
	GTTAGACTTT TTGGTGTGAA ATTAAAAGAG ATTATAGCTA AAAGTAATAA ACAAGAATAT	5220
15	ATGTATGAAC ATGCACAAGG TTTGGGTTTC AGAGGGAATA TATTTGGAGG ATATATTCTT	5280
	TATAATTTAG GGTACATTCA TTCTACAGGG CGTTACTATC TGAATGGAGA AGAAATCGAA	5340
	GACGACAAAG AATATGTACT AGGTACGATA GATATGTATA CGTTCGGTCG TTATTTCCCA	5400
20	ACATTGAAAG AATTACCAAA AGAGTATTTA ATGCCAGAGT TTTTAAGAGA TATATTTAAA	5460
	GAAAAATTAT TGGAATATTA AAAAGTAAGA TTATTGGATT TTCATTTGTC ATGAATTTTCG	5520
	ATATAATGTT TAAAGATACA CTTAACAGGA GGGTATGTGT TGTATGGCG ACAAAAACG	5580
25	AGGAAATATT ACGTAAACCG GATTGGTTGA AAATAAAATT AAATACCAAC GAAAACTATA	5640
	CAGGACTTAA GAAGATGATG AGGGAAAAAA ATCTTAATAC TGTATGTGAA GAAGCTAAAT	5700
	GTCCTAATAT ACATGAATGT TGGGGTGCAC GTCGTACAGC GACATTTATG ATTTTAGGTG	5760
30	CCGTATGTAC AAGAGCTTGT CGTTTTTGTG CGGTTAAGAC AGGTTTACCT AATGAACTTG	5820
	ATTTAAATGA GCCTGAACGT GTAGCTGAAT CAGTTGAATT AATGAATTTG AAACACGTTG	5880
35	TTATCACTGC TGTTCGCGT GATGATTTAA GAGATGCTGG TTCAAATGTT TATGCTGAGA	5940
	CAGTACGTAA AGTTAGAGAA AGAAATCCAT TTACAACGAT TGAAATTTTA CCATCAGATA	6000
	TGGGCGGGGA CTATGATGCG TTAGAAACAT TAATGGCGTC AAGACCTGAC ATTTTAAACC	6060
40	ATAATATTGA AACTGTTTGT CGCTTAACAC CGAGAGTTTC TGCGCGTGCG ACTTACGACA	6120
	GAACATTAGA GTTTTTACGT CGTTCAAAAG AATTACAACC GGATATCCCA ACTAAATCAA	6180
	GTATTATGGT TGGATTAGGT GAACTATAG AAGAAATTTA TGAAACGATG GATGATTTAC	6240
45	GTGCGAATGA TGTAGATATT TTAACGATTG GTCAATATTT ACAACCTTCA CGTAAACATT	6300
	TAAAGGTTCA AAAATATTAC ACGCCTTTAG AGTTTGGTAA ATTAAGAAAA GTGGCAATGG	6360
50	ATAAAGGGTT TAAACATTGC CAAGCTGGAC CTTTAGTACG TAGTTCTTAT CATGCGGATG	6420
	AGCAAGTAAA TGAAGCTGCT AAAGAAAAGC AACGCCAAGG TGAGGCACAG TTAAATAGTT	6480
	AATATTTAAC CATTAATAAG GCATAAAGGC TTAGTTTGTA CAAAACGAAC GTGTCATAGA	6540
55		



AGGTGAAGAA TTTGATAAAA GTAGATCAAC ATTACTTTGA ATTAATAGAA AATTATCGCG 6660  
 AATGTTTTTAA TGAAGAACAA TTTATTGCTA GGTATTCAGA TATTTTAGAT AAATATGATT 6720  
 5 ACATAGTTGG TGACTATGGT TACGATCAAT TACGATTAAA AGGTTTTTAC AAAGATTCTA 6780  
 ATAAAAAGC AGAGATGAGT AAACGTTTTT CAAATATTCA AGATTACATA TTTGAATATT 6840  
 GTAACTTTGG TTGTCCTTAC TTTGTATTAA GACATTGTG TAAACAAGAG GTTAAAAAGT 6900  
 10 TAATCGAAGA AGTTCATCCG TCTGATGTGA TAGATGACGA CAATAAACTT CAAGATGTGA 6960  
 AGATTAAGCC AACCATTCAA GATACTGAAC ATTAATAAAA CCCTTAGCTA GATTGAAAAT 7020  
 15 GGGAATCATG CAATCAAGC ATGGACCTGT AATCTAGTTA GGGGTTTTTA TCTTTAATGA 7080  
 ATGACTTCAT TTAAATACTC AGTAATTTCA TCGCCTTCTT CAGCATTAC ACCTAAAATA 7140  
 TGAGCGATAT AGCCTTCTTC TTTTAAATCA TCAGTACCGA TAATACCGAA TTTATTTGTT 7200  
 20 TGCATATTAA GTACGAGTGT CTTACCATAA TGTCTATTG TATGGACTAA CATCAAATCA 7260  
 TATCGACTAT GCTCGCCAAC AAAACCAACA AACTGAACTT GACTCTCTTC GTTGTCTATCA 7320  
 TATAAATACA TATCAATCAT TTTGTAGCGA CTCCTTTTAA AAGTAGTAAA GTTAGTATAA 7380  
 25 CGACAAATGA AGTATACTGC AAAATTATGA TAATATATAA GTGAGAGGTG ACAAGGAATG 7440  
 TATTTTGTAG ACAAAGATAA ACTAACTCAG AAATTAGCCT ATTTACAAGC ATTAAGTATG 7500  
 GATTATCATG AGAGCAAGCA CAATCAATTAT GCATTTGAAC GCATTGCTCA AATGTTGATA 7560  
 30 GAATCATCGG TAGATATAGG GAATATGATT ATCGATGCAT TTATTTTAAG GGATCCTGGT 7620  
 AATTATAAAG ATGTGATTGA TATATTAGAA CTAGAAAATG TTATTACTAA AGAAACACAG 7680  
 35 CAGGCGATTA ATAAACTGT CGGTATTCGT AAACAATTTA CATATGATTA CACAGCCTTA 7740  
 GATGTTGAGA TTATCATGCC AATGTTTGA 7769

## (2) INFORMATION FOR SEQ ID NO: 215:

- 40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 644 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

50 ACCGCCACCC ATTAATGATT GCTTAAAATC AATAGTCGTA CCATTTAATA CGGGTGCATC 60  
 TTTTTGTCT ACTAATACTT TTAATCCAAA GTATTCTAAG ACTTCATCAT TTTCACCAGG 120  
 CGCTTCTTCT GCACCCATAC CGTATGTTAA ACCAGTGCAC CCGCCACCAT TCACTTTAAT 180

55

TGCTTCTGTT AATATAACTG TTGGCATGAT AACTCCTCCT TAAAAAATCC AAGTTTCTTT 300  
 TATATGTGCA TATATATTTT GTAATAATTC TTCCGGCGAA TCACCTTCAA CAATATCACC 360  
 5 ATTTACTAAA GCATACAACC CGGCTGAACA TATACCACAA TGTGTCAGGC AACCATACTC 420  
 TAACACATCG ACATCTGGGT CATTTTCCAG TTGATTAAAA ACATAATCTC CACCTTTTGC 480  
 CATGTTAGAG AGACAAAATT CTACGATCGG ATTCATACTT CACCTTCTTA TTTCATTTGT 540  
 10 TACAATATTA TAGCATTTTA AACTGGTAT TTTAACATGA TGTGCTCAAT TAGCAACAAC 600  
 TGATGTTTCT TATCCCAGTT ATGTAATAGT GCCTTAGTTA GTAC 644

15 (2) INFORMATION FOR SEQ ID NO: 216:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1578 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

25 GAATGATGAA AGGAATAGAA AAGAAAAGAT AAATAATGTA ATAGATTTAT CCGAGAAAAT 60  
 TGAAAGAACA AAAGATATGC CAATCAAGAA TACTATAACT ACTCAATTAG GAAATAAACT 120  
 30 TATTGGCACA AAAAAAGCTC GTTTTGATGA TAAGAAAGTA GTGTCGTTTG GAGCATTTGA 180  
 AGATGAATAA AATAAATGAT AGAGATTTAA CAGAATTGAG TAGCTATAGG GTTTATCAAG 240  
 ACATCAATAA AGATAATGAC TTTACAGTTA ACGAAAAACG ATTTAAGCAG GCAGATGTAT 300  
 35 TTGAAGATTT ATATAGAGAG AACTAAAAG ACACAAATAA ATTAAGAGAG TATAATTATT 360  
 TACAAAATGA AACTTTTAAA AGCGCATAAA TAGGTGATGA GATATGCTTA AAAAAGCAAA 420  
 ATTTATCTTA ATGGCAACGA TACTACTATC AGGATGTTCA ACTACCAATA ACGAATCCAA 480  
 40 CAAAGAAACA AAATCTGTAC CAGAAGAAAT GGATGCTTCA AAATATGTAG GACAAGGATT 540  
 CCAACCACCT GCAGAAAAAG ATGCGATTGA ATTTGCAAAG AAGCATAAAG ATAAAATTGC 600  
 45 TAAGCGAGGC GAACAATTTT TTATGGATAA CTTCGGTCTA AAAGTTAAAG CTACAAATGT 660  
 TATAGGTAGT GGCGATGGTG TAGAAGTATT CGTGCAATTGT GATGACCACG AATCGTATT 720  
 TAATGCGAGT ATTCCATTTG ATAAATCAAT wATTGAsAGT GATAGCTCAT TAAGAAGTrA 780  
 50 GGAYAAAGGY GATGATATGA GTACTTTAGT TGGTGCAGTA CTCAGTGGGT TTGAATATCG 840  
 AGCACAAAAA GAAAAATATG ATAAATTATA TAAATTTTTC AAAGATAATG AAGAGAAATA 900  
 TCAATATACA GGATTTACAA AAGAAGCAAT TAATAAGACG CAAAATAGTG GTTATGAAAA 960

55

5 ACCATTGTTA AACAAAAGTG ACAGTGAATT TTCAAAAGAA TTGTCAAATG TTAAGAAGCA 1080  
 ATTAAAAGAT AAGTCTAAAG TTTCGGTAAC TACTACTCTA TTTAGTAAAA AAAAGAACTA 1140  
 TACTAAAAAA AGTAACAGTG AAAATGTAAT AAAAATGGCA GAAGAAATAA AAAAAGATAA 1200  
 AGAGATACCA AACGGTATAG AGCTTAGTAT AAAATTTTCG GACAATAAAA TAAATACGGT 1260  
 10 TAAACCAAAT TTAAACGGTG aAAGCACTTC AGAATATGGT GTGTTTGATC AAGAATAAAA 1320  
 TTAATGATGa AAATTTAACG GAGAATAGTG TATATTGAGT AGATCmAGAA TAAAAAGATA 1380  
 ATTCTACTAT TGTTGTGAAG GCAAATAAGT AGAAGATTTT AAGTGTAATT TCTGGTGATT 1440  
 15 TAAATAATAA TATAnATGGn AGTACTGATA TAAnACTTTT TAACCTACTA GATTCTTATA 1500  
 ATTTGCTTTC CATTTTATGA CGATTTTAC TCCAATTGAG TGATAGAATC CAAAAAGCC 1560  
 ATCTCCAAAA ATTAATCC 1578

(2) INFORMATION FOR SEQ ID NO: 217:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5137 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

30 TGTTTTCTT GGGTTAAAC ATGCTTGCTA TGC GTTTGTA AATATGACTT GCTGTTTTnA 60  
 CCTGnATACC CGTCACACCA TGGAAGTAAA AATGTTTCTT GCTCTGGCT TACAATTTTA 120  
 35 GCTTTAATCG CTT CATATGC TTTATATTGG TCTTCTGTTA ATTGCTGTTT TGATTCTTGT 180  
 TCGAAAACAC GATCTTTAAA TGGGTCTCTT TCAACAACCG CGTCATATTT TTCAACATAA 240  
 40 CCTTTTTGA TAAGTCCATC TAAACTGGAT TTTGAAAAGC CCATATCCTC AATATCAGTT 300  
 AAAAATATTG TTTTATGTTG TTCTTCAGAC AAGTAAGCAT ACAAATCGTA TGTTTAAATA 360  
 ACTTTCTCCA ACTTAGCTAA TACTTCATCA GGATGATACC CTTCAATGAC ACGAACAGCA 420  
 45 CGCTTGGTTT TTTTAGTTAT ATTTTGTGTG AGAATCGTTT TTTCTTCAAC GATATCATCT 480  
 TTTAACAACCT TCATAAGCAA TTGAATATCA TTATTTTTTT GCGCATCTTT ATAATAATAG 540  
 TAACCATGCT TATCAAATTT TTGTAATAAA GCTGAAGGTA GCTCTATGTC ATCTTTCATC 600  
 50 TTAAATGCTT TTTTATACTT CGCTTAAATA GCACTCGGAA GCATCACTTC TAGCATAGAA 660  
 ATACGTTTAA TGACATGAGT TGAACCCATC CACTCACTTA AAGCTATTAA TTCTGATGTT 720  
 55 AATTCTGGTT GTATATCTTT CACTTCTATG ATTTTTTTTA ACTTCGAAAC GTCAAGTTGT 780

	ACAATTACAC GCACACCAGG TTGGATGACA GATTTCGAGTT GTTCGGGAAT TATATAATCA	900
	AATTTATAGT CAACGCTCTT CGACGCGACA TCGACTATGA CTTTCGCTAT CATTATTGCC	960
5	ACCTAGTTTC TAGTTCATCT AAAATTTGTG CAGCTAATAC TACTTTTTTT CCTTCTTGA	1020
	TATTTACTTT TTCATTATTT TTAAAATGCA TTGTCAATTC ATTATCATCA GAACTAAATC	1080
10	CGATAGACAT ATCCCCAACA TTATTTGAAA TAATCACATC TGCATTTTTC TTGCGTAATT	1140
	TTTGTTGTGC ATAATTTTCA ATATCTTCAG TCTCTGCTGC AAAGCCTATT AAATACTGTG	1200
	ATGTTTTATG TTCACCTAAA TATTTAAGAA TGTCTTTAGT ACGTTTAAAA GATACTGACA	1260
15	AATCACCATC CTGCTTTTTT ATCTTATGTT CTAATACATC AACCGGTGTA TAGTCAGATA	1320
	CGGCTGCTGC TTTTACAACA ATATCTTGTT CGTCAAATCG GCTTGTCAC TGTTCAAACA	1380
	TTTCTTCAGC ACTTTGAACA TGAATAACTT CAATATCTTT TGGATCCTCT AGTGTGTAG	1440
20	GACCAGCAAC TAACGTCACG ATAGCTCCTC GATTTTCGCA TGCTTCAGCT ATTGCATAGC	1500
	CCATTTTTCC AGAAGAACGA TTGGATACAA ATCTGACTGG ATCGATAACT TCAATAGTTG	1560
25	GTCCTGCTGT AACCAATGCG CGTTTATCTT GAAATGAACT ATTAGCTAAA CGATTACTAT	1620
	TTTGAAAATG AGCATCAATT ACAGAAACGA TTTGAAGCGG TTCTTCCATA CGTCCTTTAG	1680
	CAACATAACC ACATGCTAGA AATCCGCTTC CTGGTTCGAT AAAATGATAC CCATCTTCTT	1740
30	TTAAAATATT AATATTTTGC TCGGTACGTT TATTTTCATA CATATGCACA TTCATAGCAG	1800
	GCGCAATAAA TTTCCGTGTC TCTGTTGCTA GCAACGTTGA TGTCAACCAA TCATCAGCAA	1860
	TACCTACACT CAATTTTGCA ATTGTATTTG CCGTTGCAGG TGCAACAATG ATTGCATCTG	1920
35	CCCAATCACC TAATGCAATA TGCTGTATTT CTGAAGGATT TTCTTCTATA AAAGTATCTG	1980
	TATAAACAGC ATTTTCGACTT ATTGCTTGAA ATGCTAATGG TGTCACAAAT TTTTGTGCGT	2040
40	GATTGTTTAA CATAACGCGA ACTTCATACC CAGATTGTGT TAACTTACTT GTCAAATCAA	2100
	TTGCTTTATA TGCCGCAATG CCACCTGTAA CGGCTAATAA TATTTTCTTC ATATTCAATC	2160
	TCCCTTAAAT ATCACTATGA CATTTACGCT TTACATCATC ATATGCGCAC AAATGCTCAT	2220
45	TACTTTTTTA TAGATACAAA TTTAGTATTA TTATAACATC AATCATTGGA TAAACTAAAA	2280
	AAACACACCT ACATAGGTGC GTTTGATTTG GATATGCCTT GACGTATTTG ATGTACGTCT	2340
	AGCTTCACAT ATTTTAAATG GTCGAAACTA TTCTTTACCA TAATAATCAC TTGAAATAAC	2400
50	AGGGCGAATT TTACCGTCAG CAATTTCTTC TAACGCTCTA CCAACTGGTT TAAATGAATG	2460
	ATATTCACCT AATAATTCAG TTTCAGGTTG TTCATCAATT TCACGCGCTC TTTTCGCTGC	2520
55	AGTTGTTGCA ATTAAATACT TTGATTTAAT TTGTGACGTT aATTGGTTcA AAgGTGGATT	2580

	TTTA <del>r</del> GTGcT	CAGCTTCTAC	AATACATTGA	ATTc <del>t</del> ATTcy	TCGcAAGtTC	TACTTCatCA	2700
	TTAACTACAA	cGTAAyCGTA	TAAATTCATC	ATTTCrACTT	CTkTACGCGC	yTCGTTAATA	2760
5	CGACTTTGTA	TTTTCTCATC	AGATTCTGTT	CCTCTACCTA	CTAATCGCTC	TCTCAAGTGT	2820
	TCTAAACTTG	GAGGTGCTAA	GAAAATAAAT	AGCGCATCTG	GAAATTTCTT	TCTAACTTGC	2880
10	TTTGACCTT	CTACTTCAAT	TTCTAAAAAT	ACATCATGAC	CTtCGTCCAT	TGTATCTTTA	2940
	ACATATTGAA	CTGGTGTAAC	ATAATAGTTG	CCTACATATT	CAGCATATTC	TATAAATTGG	3000
	TCATCTTTGA	TTAAAGCTTC	AAACGCATCC	CTAGTTTTAA	AAAAGTAATC	TACGCCATCA	3060
15	ACTTCACCTT	CACGCATTTG	ACGTGTTGTC	ATTGAAATAG	AATACTTATA	TGATGTACTT	3120
	GGATCTTCAA	ATATnCGTnT	TCTAACAGTA	CCTTTACCTA	CTCCAGATGG	TCCTGATAAA	3180
	ACGATTAACA	ATCCTTTTTC	ATTATCCATG	CCTTACGACC	TCTCTAAGCT	AATCTTCTAT	3240
20	TATTTAAATA	TGATATCACA	TTGTTCTTTA	TATTGTATAG	CATATTTGAA	ATTGCATGCC	3300
	ATAATTTCTA	TTAAGTCTAA	CAATATCGTT	ATATTGCACG	ATTAATTTTA	ATTAAATAAA	3360
25	TTGAATTGCA	AACTTTTAGA	TAATGTAAAA	TGTATGGCAT	AATGTATGGT	TCAATAACTA	3420
	TACTGAAAAG	TTACAATCAT	GTTAAAATGA	AACGAATGAT	ATGAAGAAGG	TGGAAGATAA	3480
	ATTATGGCTT	ATGATGGCTT	ATTTACAAAG	AAAATGGTTG	AGTCTCTACA	ATTTTTAACA	3540
30	ACAGGACGTG	TTCACAAAAT	CAATCAACCT	GATAATGACA	CGATACTAAT	GGTGTACGT	3600
	CAAAATAGAC	AAAACCATCA	ATTGTTATTG	TCAATCCATC	CAAACTTTTC	AAGATTACAA	3660
	TTGACTACTA	AAAAATATGA	TAATCCATTT	AATCCACCCA	TGTTTGCGCG	TGTTTTTAGA	3720
35	AAACACTTAG	AAGGTGGTAT	TATCGAATCG	ATTAAGCAAA	TTGGTAATGA	TCGTCGCATT	3780
	GAAATCGATA	TAAAGAGTAA	AGATGAAATT	GGCGATACTA	TTTACCGCAC	TGTCATCCTT	3840
	GAGAT <sup>5</sup> TATGG	GTAAACATAG	TAACTTAATT	TTAGTAGATG	AAAATCGCAA	AATAATTGAA	3900
40	GGATTTAAAC	ACTTAACACC	AAATACGAAT	CACTATCGTA	CAGTAATGCC	AGGATTTAAT	3960
	TATGAAGCAC	CACCTACTCA	GCACAAAATA	AATCCGTATG	ATATTACAGG	TGCAGAGGTG	4020
45	TTGAAATATA	TCGATTTTTAA	CGCAGGTAAT	ATTGCTAAAC	AATTATTGAA	TCAGTTTGAA	4080
	GGATTTAGCC	CTTTAATTAC	GAATGAAATC	GTTAGTCGTC	GTCAATTTAT	GACTTCATCA	4140
	ACATTACCAG	AAGCATTTGA	CGAAGTAATG	GCAGAAACCA	AGTTACCACC	TACTCCTATT	4200
50	TTTCATAAAA	ATCATGAAAC	AGGTAAAGAG	GATTTCTATT	TTATAAAGTT	AAATCAATTT	4260
	AATGATGATA	CAGTTACATA	CGATTCATTA	AATGATTTGC	TTGATCGTTT	TTATGATGCG	4320
	CGTGGCGAAC	GTGAACGCGT	TAAACAACGT	GCGAATGATT	TAGTTCGATT	TGTTCAACAG	4380
55							

5 ATAAAGATAC TGAACAGTTA TATGGTGAAT TGATCACTGC TAATATATAT CGAATTAAGC 4500  
 AAGGCGATAA AGAAGTGACG GCATTGAATT ATTATACGAA TGAAGAAGTT GTCATTCCTT 4560  
 TAAATCCTAC AAAATCCCCA TCAGCAAATG CTCAATATTA TTATAAACAA TATAAYCGTA 4620  
 TGAAAACGAG AGAmCGTGAA TTACAACATC AAATTCAATT GACGAAAGAC AATATAGATT 4680  
 10 ATTTTTC AAC AATCGAACAA CAATTACATC ATATTTCTGT CCATGACATT GATGAAATTA 4740  
 GAGATGAATT AGCAGAACAA GGCTTTATGA AACAGCGTAA AAATCAAAC T AAGAAAAAGA 4800  
 AAGCGCAGAT TCAATTACAA CATTATGTAT CAACTGATGG CGACGATATA TATGTTGGTA 4860  
 15 AGAATAACAA GCAAAATGAT TATTTAACAA ATAAAAAGC TAAAAAACT CACACATGGT 4920  
 tACACACAAA AGATATTCCT GGTTACATG TCGTTATATT TAATGATGCA CCAAGTGATA 4980  
 CGACAATCAA GGAAGCGGCT ATGTTAGCAG GATACTTTTC AAAAGCTGGT AATTCTGGAC 5040  
 20 AAATACCTGT TGATTATACA TTAATTAAAA ATGTGCATAA ACCATCaGGT GCAAAGCCTG 5100  
 GGTTTGTAAC ATATGACAAT CAAAAAACTT TGTATGC 5137

25 (2) INFORMATION FOR SEQ ID NO: 218:

. (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2267 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:  
 35 GTTTTATCGC AGCAGTAAAG CTATCAATCG GCGGTTCAAT TGATGATGCA TTAGCAGAAA 60  
 TCAaACAATC ATTTTAGTTA AAATTTACTA ATAATGAaAA ATGTAAACCT TTTTCAAATG 120  
 AAAC<sup>~</sup>TTTATa AaAAATATGA TAGTATATAT GTAAATGTTT AATAAAATCT GGAGAAATAG 180  
 40 GAGGACATTG CCATGCAACA CCTTATAAAA AAACATGTAT TGAATGGCGA GTTTGATTTA 240  
 GTACGACAAT TGATGTCCGA AACAGATTTT ATGGAATTTG AAGAAGCATA TATTTCAAGT 300  
 GCGCATGAAG TAGAAAGTAT GATGTTTTAT ACATGTATTT TAGATATGAT TAAGTACGAA 360  
 45 GAATCATCTG AAATGCATGA CTTAGCATTT TTATTGCTTG TGTATCCACT AAGTGAATAT 420  
 GAAGGTGCTT TGGATTCTGC TTATTATCAT GCAGACGCTT CCATAAACT TACTGACGGC 480  
 50 AAAGAAGTTA AAAGTTTGTT ACAAATGTTA TTATTGCATG CGATACCAAC ACCTGTTATT 540  
 TCAGATAAGA AGGCTTTTGA TATCGCCAAG CAAATTTTAA AATTAGATCC TAATAATAAT 600  
 GTTGCTCGTA ACGTCTTAAA AGACACTGCC AAACGTATGc gACaACGTTG TTGTTGATAT 660  
 55

AGTTTAAACA TTTGGTTGGG TTGGGCATAT GTTCCAGCCT TTTTAAATAC TTAAAACTA 780  
 ACGAAgTATA CTTGTGTGCA CAAATGGTTT TTATACAACA TTTTATAAAT TTATACATTT 840  
 5 TAATAAAGAA CATACGATAG ATGGTTTAAA CCTTGTTAAC TGAGAAATTT TGATATGTAT 900  
 TCTTCGAAAT TTAActAAAT ATACGAAATT CAAGAAGCAC AATAATTAAT CATTTTTCCT 960  
 10 ATACAAAAGT TCGTATGACT GCATTATAAA AGCATAAATT TATAATTTTT TTAAATGTCA 1020  
 TTGAACGTGA TAATGTGAAT GGATTGAGCA ATTTTGAAAA AGTGAAAAAT AACCTATGCG 1080  
 ACTTGCAATT AATTTTCAGT ACGTTATAAT GCACACTGTG CAAAATTAAG GAGGTCTATT 1140  
 15 ATTCACATGA TGATGAaTAA AGAAGCAACA AAAATTGGAT TTGCCTACGT CGGCATTGTA 1200  
 GTGGGCGCAG gATTTTCAAC TGGACAAGAA GTTATGCAAT TTTTCACTAA ATATGGCTTG 1260  
 TGGGCTTATT TAGGTGTTAT TATATCTGGT TTTATTTTAG CTTTATTGG GCGCCAAGTA 1320  
 20 GCAAAAATTG GTACTGCCTT TGAAGCGACA AATCATGAAT CAACATTACA ATACGTATTC 1380  
 GGTGAAAAGT TTAGTAAAGT CTTTGaTTAT ATTTAATCT TCTTCTTATT TGGTATAGCT 1440  
 25 GTAACCATGC tAGCTGGTGC AGGCGCAACA TTTGAAGAAA GTTATAACAT ACCTACATGG 1500  
 CTAGGTGCTT TaATTATGaC ATTAGCGATT TATATTACGT TGCKATTAGA CTTTAATAAA 1560  
 ATAGTACGTG CACTAGGTAT CGTTACACCA TTTTAAATTG TTTTAGTTGT ATTAATCGCT 1620  
 30 GGCGTTTATT tATTTAAAGG TCATGtTTCA TTAGCAGAAG TTAACCAAGT AGTGCctGAA 1680  
 GCAAGTATTT GGAAGGGAAT CTGGTTTGGT ACAATATATG GTGGATTAGC TTTTCTGTA 1740  
 GGTTTTAGTA CCATCGTAGC AATCnGTGGG GATACTGAAA AGCGTACAGT GTCAGGTGCA 1800  
 35 GGCGCGATGT ATGGTGGTAT TATCTATACT GTATTACTAG CATGATCAA CTTTGcATTG 1860  
 CAAGTGaATA TCCAActATT AAAAATGCCT CAATTCCTAC ATTGACGTTA GCAAATAATA 1920  
 TCCA<sup>1</sup>CCCTT AATAGCAACA GTGktATCTG TTATTATGCT GGCGGktATG TATAATACTA 1980  
 40 TTCTAGGACT AATGTATTCA TTTGCAGCAC GTTTTACAGA ACCATACAGT AAAAATTATC 2040  
 ATATCTTtAT TATTATAATG ATGGTAGCAG GTTATTTATT AAGTtnCGTA GGATTtGCTG 2100  
 45 AATTAATTAA TAAGTTATAT ACnATTtATG GGATATGTAG GCTTATTnTA TTGTAGTAGC 2160  
 TGTAATTATn AAATATTTCC AAACGTAAAA ATGGCGGATA AAAACATAT TGCTTTAATA 2220  
 TCATATGGAG GGGATATCCG AAACTTTACA ATTTGAATCA CTTTGGT 2267

50 (2) INFORMATION FOR SEQ ID NO: 219:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6336 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

5	GTCAATGTAA CCTAATAGTT TATGTCTATC TTGTGTACCA ACTACTACAT CGACACCAGG	60
	AATTTCCATA ATTTTCAGCTG ATGAAGTTTG CGCATAACAA CCTGTTACAC AGATTACAGC	120
10	ATCAGGATTT TGTCTTATTG CACGTCTAAT TATTTGACGA CTTTTTTTAT CACCCGTATT	180
	CGTTACTGTA CAAGTATTAA TAACAAATAC ATCAGCATTG GCTTCAAAGT CAACGCGCTC	240
	ATAGTTTGCT TCTTTAAATA ATTGCCAGAT TGCTTCAGTT TCATAATGGT TTACTTTTACA	300
15	ACCTAATGTG TGaACGCAAC TGTTGACATA AATATTCACC CCATTAATTC TTTTTCATAA	360
	CTTATTGCAC TTAACGCATA CAATGGCGCA GTTTCTGCCC GTAAAATTCT CGGCCCAAGA	420
	CCAACAACCTG TACTAGTATT ACTAAATAAT GAAATTTTCT TTTCTGACAA ACCACCCTCA	480
20	GGaCAAAAA TCATCAACAC TTTATCCTGA GCATTGAATT GTTGTAAGT TTGCTTGAAA	540
	TTGCTTAACT CACCATCTTT TGCTTCCTCT TCATATGCAA TAAGAATATA GTCATAATTA	600
	TCAATAGTAT CACAAATTAA TTTTAAATTC GACTCGAATT GAATAGATGG AATCACTAAA	660
25	CGATAGCTTT GTTCAGCAGC TTCTTTAATT ATTTTTTGCC AACGCTCTAT CTTTTTGGCA	720
	ACTTTTGCCT CGTTTAATTT AACAATTGAA CGTTCATGC TCACAGCTAT AAATGATGAA	780
30	GCACCCAATT CAGTAGCTTT TTGTAGCAAC CACTCATATT TGTCAGCTTT GATTAGTCCA	840
	CTGCAAATCG TAACATCAAC TGGCAATTCT GTATTAATAT TTTGTTTTTC TTTTAAATCA	900
	ACTTCAATTT TATCACTTGT TATGTCAGCA ATTTACATA AATAAACTGT TTGATCATTA	960
35	AAAGTTAAAA TAATTTTACT ACCAACATCA TATCTCATTA CATTTGTTAT ATGATGAATA	1020
	TCTTCTTTTT TTGTAATAAA AAAACGCTGA CTTACATCAG CGTTTTGGnT CTATGAAATA	1080
	ACGTTGCACA TTATTCATC ACTTTCTGGC CAACAAGACA AACCCAACCG TTGTCATGTT	1140
40	GTCTGAAAT AATTTTAAAA CCTACACGCT CCATATGTGA CTGTATACCT TCATACTTCT	1200
	CTTTTATAAT ACCAGAAGTA ATAAAATAAC CGCCTTCATT TAGAGTATTA TAAGCATCTT	1260
45	CAATCATTTT ATCAATAATA TGCCTAAAA TATTTGCTAT TACAATATCA AATTTTTCTG	1320
	TTTCGTCTTT CAATAAGTTA CCTGGAACAG CTTCAATTAA CGTTTCACAA TGATTTCTTC	1380
	TGAAGTTTTT TTTAGCTACA CTCCTGCCA TTTCATCAAT ATCCAACGCT TTAATACGTT	1440
50	TTACACCGAT TAGATGACTT GCAATACTTA ATATACCTGA GCCAGTACCA ACATCAATTA	1500
	CTGAATGCTG TGGCAATACA TATGTTTCTA TTGCCTTCAA ACACATACTT GTAGTCGGAT	1560
55	GATCACCTGT TCCAAAAGCC ATACCTGGGT CGAGCTCAAT GCAAAGCTCT TCATCCGCTT	1620



	GGAAATAGTT TTTCCATTCA TTTTCCCAAT CCGTCTCTGC AATAATTTGC TCACTGAATT	1740
	GAACGTTATG TTGATCAAGT TCATCTAAAT TTAATAACTC ATCTTTAATT TGCTGTCGCA	1800
5	ACTTATCATC ATAAGTCATT TCATTAAAAT AGGCTTTCAA TCTTACTCCC TTATCTGGAT	1860
	AATCCTCTTT TTTCAAAGCG TAAATTTTAC CGTATTTATC TTCTGGTTGG TTAATTAAAT	1920
10	CATCTGAATC TTCTATCAG ACACCATTTG ATCCATGATT TTCAAGTATA TTGGTAGCCA	1980
	ATTCTACTGC TTCATGATTA ATAATAATTG AAAGCTCTGT CCAGTTCATA CTTTATTCTC	2040
	CCTTAAAGAA TCITTTTGCT CTATCTTTAA AATTCGAAGG TTGTTTATTA ATTTCTTCAC	2100
15	CATTTAATTG GGCAAATTCT TTCATTAGTT CTTTTTGTCT ATCTGTTAAT TTAGTAGGCG	2160
	TTACTACTTT AATATCAACA TATAAATCTC CGTATCCATA GCCATGAACA TTTTTTATAC	2220
	CCTTTTCTTT TAAGCGGAAT TGCTTACCTG TTTGTGTACC AGCAGGGATT GTTAACATAA	2280
20	CTTCATTATT TAATGTTGGT ATTTTATTTT CATCGCCTAA AGCTGCTTGT GGGAAGCTAA	2340
	CATTTAATTT GTAATAAATA TCATCACCAT CACGTTTAAA TGTTTCAGAT GGTTTAACTC	2400
	TAAATACTAC GTATAAATCA CCAGCAGGTC CTCCATTAC GCCTGGAGAG CCTTCACCAG	2460
25	CTAATCTAAT TTGTTGTTCA TTGTCGACAC CTTCAGGTAC TTTCACTTCT AATTAACTG	2520
	TTTTATTTTC AGTACCTTTT CCGTGACATG TTGGACAAGC TTCTTCAAAT TCTTGACCAC	2580
30	TTCCATTACA TTTAGGACAA ACTTGTTTCTAG TACGAACTCT ACCTAAAATT GTGTTTGTGTT	2640
	CTACAGCTAC ATGACCAGCG CCATTACAGT AACTACAAGT CTTTTTACTT GTTCCAGGCT	2700
	TTGCACCATC ACCATGACAT GTTTCGCATG TTACATCTTT ACGGATTGAA ATTTCTTTTG	2760
35	TTGTACCAA TACCGCTTCT TCAAATGTTA ATGTCATTGT AACTGAAGA TCATCACCTT	2820
	TTTGCGGTGC ATTTGGATCT CTTTGTCTGC CGCCACCGAA GAAAGAGCTA AAGATATCTT	2880
	CAAA <del>5</del> CCGCC GCCACCGAAG CCACTAAAAC CGCCAAAGTC AGAGCCATTG AATCCTTGTC	2940
40	CACCAAACC TTGTGGACCA TCATGTCCAA ATTGATCATA GcTTGCGCGT TTATTATCAT	3000
	CACTTAAAAC TTCATAGGCT TCAGAAATTT CTTTAACTT TTCATCTGCA CCTTCTTCTT	3060
45	TGTTAATATC TGGATGATAT TTTTTCGAAA GCTTTCGATA CGCTTTTTTG ATTTTATCTT	3120
	TTGAAGCATC CTTACTAATG CCTAAAACCT CATAATAATC TCITTTGGCC ACAGCTATCT	3180
	CTCCTTTTCT TAATTAACTC ATATAGTTTA ACGTAATATG TCATACTATC CAAATAAAAA	3240
50	GCCAAAGCCA ATGTTCTATT GACTTTGACT TTTTCTGATCA TGACAACATT CTAATTGTAT	3300
	TGTTTAAATTA TTTTGTGTCG TCGTCTTTTA CTTCTTTAAA TTCAGCATCT TCTACAGTAC	3360
55	TATCATTGTT TTGACCAGCA TTAGCACCTT GTGCTTGTTG TTGCTGTTGA GCCGCTTGCT	3420

	TATCTTCTAT ATCTTGACCT TCTAAAGCAG TTTTAAGAGC GTCTTTTTC TCTTCAGCAG	3540
5	ATTTTTTATC TTCTTCACCG ATATTTTCGC CTAAATCAGT TAAAGTTTT TCAACTTGGA	3600
	ATACTAGACT GTCAGCTTCG TTTCTTAAGT CTACTTCTTC ACGACGTTTT TTATCTGCTT	3660
	CAGCGTTAAC TTCAGCATCT TTTACCATAC GGTCGATTTC TTCGTCTGAT AATGAAGAAC	3720
10	TTGATTGAAT TGTAATTCTT TGTCTTTTAT TTGTACCTAA GTCTTTTGCA GTTACATTTA	3780
	CAATACCGTT TTTATCGATA TCAAACGTTA CTTCAATTG AGGTTTACCA CGTTCAGCTG	3840
	GTGGAATATC AGTCAATTGG AATCTACCAA GTGTTTTATT ATCCGCAGCC ATTGGACGTT	3900
15	CACCTTGTA TACGTGTACA TCTACTGATG GTTGATTATC TACTGCTGTT GAATAGATT	3960
	GAGATTTAGA TGTAGGAATC GTAGTGTTAC GTTCAATTAA CGTATTCATA CGTCCACCTA	4020
	AAATTTCAAT ACCTAAAGAT AGTGGTGTTA CGTCTAATAA TACTACGTCT TTAACGTCAC	4080
20	CTGTGATAAC GCCACCTTGG ATTGCAGCTC CCATTGCCAC TACTTCGTCC GGGTTTACTC	4140
	CTTTGTTAGG CTCTTTACCG ATTTCTTTTT TGACAGCTTC TTGTACTGCT GGAATACGAG	4200
25	TTGATCCACC AACTAAGATA ACTTCATCGA TATCTGAGTT TGTTAAGCCA GCGTCTTTCA	4260
	TTGCTTGGCG TGTAGGTTCC ATTGTTCTTC TAATTAATGA ATCTGATAAT TCTTCAAATT	4320
	TAGAACGAGT TAAGTTTACT TCTAAGTGTA ATGGACCGTT TTCACCAGCT GAGATAAATG	4380
30	GTAATGAGAT TTGAGTTTGT GATACACCTG ATAAGTCTTT TTTAGCTTTT TCAGCAGCAT	4440
	CTTTCAAACG TTGTAATGCC ATTTTATCTT GAGATAAGTC TACGCCATTT TCTTTTTTGA	4500
	ATTCTGCAAC TAGGTAGTCA ATAATTACTT GGTCAAAATC ATCACCGCCA AGTTTGTGT	4560
35	CACCGGCTGT TGATAGTACT TCGAATACAC CGTCACCTAA TTCTAGGATA GATACGTCAA	4620
	ATGTACCGCC ACCTAAGTCA AAAACAAGAA CTTTTTCATC TTTATCAGTT TTGTCTAAAC	4680
40	CATAFGCTAA TGCTGCAGCT GTTGTTTCAT TAATGATACG CTCAACTTCT AAACCAGCAA	4740
	TTTTACCAGC ATCTTTAGTT GCTTGACGTT CAGCATCGTT AAAGTATGCA GGTACTGTAA	4800
	TTACAGCTTT GTCAACTTTC TCACCTAAaA TAgTTTCAGC TGTATTTTTT AAGTTTTGTA	4860
45	AAATCATAGC TGAGATTTCT TGTGGTGTGT ATGATTTACC TTCAATATCT ACTTTATAAT	4920
	CAGTACCCAT ATGACGTTTA ATAGATTGAA CAGTGTTTGG GTTTGTAATA GCTTGACGTT	4980
	TTGCTACTTC aCCAACTTGA GTTTCTCCAT TTTTGAAAGC TACAACAGAT GGTGTTGTAC	5040
50	GTGAACcTTC AGGGTTTTGA ATTACTTTTG GCTCATCGCC TTCTAATAcT GThACACATG	5100
	AATTTGTTGT ACCTAAGTCT ATACCAATAA TTTTACTCAT AATAAAATTC CTCCATTTAA	5160
55	TCATTAAATT AATTTAATTT TAAACAATGT CTTTTCGCCA AATTTAAGTT ATTGGTTTAC	5220

AGTGATTTCG CCAGATTCAA AATCAGGGTT ATCATCTTGA ACTACAGCTT GGTGAATATT 5340  
 TGGATCAAAT GCTTCACCTT CAGTTTAAAT AACTTCAAGA CCATTATCTT TTAGTGCGTT 5400  
 5 AATCAAACCTT TCATGCACCA TTTGTACACC TTTTGAAGA GATTTAAAAG TCTCATCATC 5460  
 ACCTTCAATT TGAAGTGCAC GTTCTATATT GTCTATTGCT GGTAATAATAT CTGTAAACAC 5520  
 10 ACGTTGTGCT TGATATGTTT TGTTTATTTT ATTTTCTTTT TGAATTCTAC GCTTATAATT 5580  
 TTCAAACCTCA GCGTAGAGCC TTAAATATTT CTCTTCGTTT TCATCTGCTA ATTGTTGAAG 5640  
 TTCATTAATT TTTTGATCTT TTGGATCTAT TTCTTCAATA ACATTCTCGT CAGACGTTTC 5700  
 15 TTCTATTGCT TCATCTTGTA AATGACCTTT ACTTTCTTCA GCTTGTTCOA CTGAATCATC 5760  
 AATATTTTGT TTGACGTTTG TTTCTTCAAC TGTTGATTCA GTGTTTTTTT CAACTGATTC 5820  
 GTCTTTATTT GTCATTTTCT GTCCTCCAAT ACTTTCTAAT CCATCATTAC CAAATTCTAT 5880  
 20 TTAATAATTG AATGACATTT TGATAATGCA TAGCTGTAGG TCCAATCACA GCGATTTGAC 5940  
 CTTTTAACGT TTCATCAAAA TGATATTGAC TTGTTACAAT TGAAATATCA CTTAAGCTGT 6000  
 CATCAATTTT ATTACCAATT TTTACATTAA TATTTGGTGA AGATATATCT TGTAATAATT 6060  
 25 CTGCAATTCT ATTTGATTCT ATATATTGTA GAATGGGCTG AATTGAAGAT ACATTACTTT 6120  
 CATTCAATGC ATCAATAAGT TTAACCTTTC CACCCATATA AATGCTATTA CTTTGATTAG 6180  
 30 AAATATGATT ATTCATCGTA TTTAACAATT TATTGATAAA AATTTCTTCC TGCTCTGATT 6240  
 GAaCAAAAGA GACAATATCA TCTTGTAAT TCTGATTAAA CTCAGTTAGT TTGTTTGTA 6300  
 CAAAATTGTA TATTGTATTT AGTTTGTTCAT TATTAA 6336

35 (2) INFORMATION FOR SEQ ID NO: 220:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13059 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

TTCATGATTA TTATCTGTTG TAGACACTGC TGGATCTTCC GATGTATCTT TCGATGCATC 60  
 TTTGATTTG TGTATTTGCT GATTCAAATG GTCTAGGTCT TCTAACGCCT TATTTACCAT 120  
 50 TGCTTCATCA TTTTATCAT CTTTTTCTCC ATGTTTTGTT GTAGCCGTTT GTGACATATC 180  
 ATTTTTTATT GCATTAAGAT CGTCCTCGCC ACTTTGTTGA CCCCTATCAA CATTGAAGA 240  
 AACCTCATTT AAATCTTTAA GCAATTGATC TAATTTACTG TCTATATCAC TTTGACCGTT 300

55

	TTCATCTATT	TGCGATGCTG	TTTTCGCTTC	ATTTAGTTGT	GCTTTATAAT	GTTCTTTAGA	420
5	TGAAGCCGAT	AACTGTTTTA	ATTGCTCAAT	TTGACGAATT	GCCTTGTCAA	CTTTGTCTAA	480
	TAAATCTTGC	TTAGATAATA	TCTCTTTTGT	AATTTTCAGTA	TCCTTTTCAG	ATGCAGCTTG	540
	GGCATCGTAC	GGCAAGATAT	TCGTTAAAAT	GATACTTGTC	GCCATCATTG	TCGAACACGA	600
10	TAACTTTACA	TATAATTGAA	ACGGTTTCCC	TCGATATTTA	GCCATCAACA	TACTCCTTCC	660
	TCACTTACTT	CCTTCAAAGA	ATTACATACT	ATTATATAACC	TGTTTACAAG	AAATTTACAC	720
	TTATCTATCT	AGTTATTGTT	GTTAGTAATT	ATCTACTTAT	TACTTAGCTT	ATATTTAAGT	780
15	AAACAAAACA	AGCATGACGT	AATATCATAT	TGTCCATGTC	GCTAACATCA	TATTACGTCA	840
	AATCTTTTAT	ATTAAATGAT	GTTTTATTTT	AGACTGCTTT	TTCTTTTtag	CTTTCGAGCG	900
	CCTGTTTAAA	AACTTGCTCG	AATTGTTTAC	GCGAGATTTT	GTGTGCATGT	GCTTTTTGTG	960
20	CTAATAAAGC	ATCTCGAAAC	TGTTGTTGAT	CTTTCAAACT	TTCTAACATT	TGTATTAATT	1020
	GGTCTTTACT	TTCCATTGTT	ATCTCATCAT	TATGCTCAAA	TAAGTGCTCT	GATAATGTTA	1080
25	CTTTAGCATG	GTGTGCGGTT	TGACGATAAC	CTAAAATCAA	CAACTCATAG	TCAAACGCTT	1140
	GTTCCACCGC	ATTTAAAATT	TCATTACCCT	CATTGATATC	AAGATAAATA	TCACATAACT	1200
	GGTATAGTTC	ATTTACCCTG	TCAATATTAA	TAGATGGGTA	TAAATGCACA	TTAGCATATT	1260
30	GATCAAGTTG	CATTAGCTTA	TCAGACATCT	CTGTAATAGC	AGCGATGTGn	AACTTAAAAT	1320
	CTGGTAAAGt	TyCAACCAAT	ACCTTGATGT	TACGAatTGa	TCCgAGTTAG	TTAATATTAC	1380
	AATTTCTTTA	GTATATCTAT	TACGACTACG	ATAGTTATAT	AGATATCCGC	CTTGTAATAAT	1440
35	ACGAGATTGA	ACCTTTGCGT	CTGCTATATT	GAGCATCGTT	TCATATTCGT	TTTATCTGG	1500
	AATAATAATA	TTACAATGTC	GTTTCATATC	ACCTTTACAC	ATCAATTGCA	TATTTCCCGG	1560
40	GACATTACCA	TTACAGTGTT	CTTGCCATAC	CAAAACATCA	CTACCTTTTG	ATGGCAAATT	1620
	ATATAACACT	GAAAATGGTA	GGGCTAGTGA	GTTAATAACG	AAATGATGTT	CCGTAATTTT	1680
	AAGTTGCTTG	ATAAAAAATA	ATGCGAATGC	GAGCTTTGAA	GGGAAAAAGT	AAGACTTCCC	1740
45	TTGCCAATCC	AATATGACAT	CAGATGTTAC	AAAATTTTCA	TAAATCACTT	CTTTACCTTC	1800
	TGCTGTCATA	TATTTCTTCA	AGATCGCTTT	ACGATTTAAA	TCGTAAACAG	TTTGTCAAA	1860
	TTTAATACCA	TTCTTAGAAT	AATAATCGAC	AAATCGGACA	CGTTGTTGGT	CATCAAACCA	1920
50	TTGACACGA	CTAACAATTC	TAGGGCGCTC	TCCACTTTGA	yAAAATATTT	TACCTCGyAG	1980
	ACGTCCCATA	TCATTaATTG	TAGCCGAATT	GTTGTTACCT	TTAATTTCCC	AAAAAGCTGG	2040
55	TACAGTAACC	TGATTAAAAA	ATCGTGTTTT	CATATTTTCT	GTATTATGAT	TATCTGCAAA	2100

	TAAATCTTCT	TCCAACCTTAC	TGGCTTTAAA	AGACTCATAT	AACTTTCGTG	AATGATCGTT	2220
5	AAAGTAATCA	AATAATTTAA	TCATGTAGCA	CCTCTTGaAC	TAATGTTTCC	CATTTTAAAA	2280
	TAATATCTTG	AGTCATAAAT	TGCTGTGCCA	CTTCATAAGA	GATGTCATGT	GGTGTCTGGG	2340
	GACCATTGTT	AAAATACATT	ACAATGGcAT	GAGCTAGTTT	TGCGATAACA	TCATCCACAC	2400
10	TATCTTCGTC	GGTATCAAAA	GGTACCAAGT	AGCCATTTTC	CCCATCTCGA	ATAAAGGTTG	2460
	GGTTACCATA	ATTCACATTT	AATCCAATCA	TACCTAGTCC	TGAGCCTACC	GCTTCCATTA	2520
	GTGTTAACCC	AAAACCTTCG	CTAGTTGATG	CAGAAAGAAA	TAACTCATAA	TCATTATAAA	2580
15	TTTCATCAAG	TTTAACATGC	CCTAGTAAAC	GAATATAATC	TTGTGCGCGG	TGTGTATCAA	2640
	TAATTTTACG	CAGTCGCGTC	TTCTCGCTAC	CTTCTCCATA	AATATCAAAT	GTTAATTCTG	2700
	GCACTTGTCG	TTTAGCCACG	ATAACCGCTT	TGACAAGCCA	ATCAATATGT	TTCTCATTCG	2760
20	CTAAACGAGA	TGCACTAATC	ATCGCATATG	GCTTTCTTGA	TAATGTTGGA	TATGATAATG	2820
	CATCAATGCT	TCCCACAGGA	ATAGTATAAA	CACGTGGGCG	ATAACCTTGA	TATTGCTCAA	2880
25	ATTGTCGACA	AACCATATGA	TTTTGAATAT	CTGTTGCTGT	AATAAAGAAA	TCAATGTATT	2940
	TAGCTTTTGA	AAATTGATAT	TCATAATAAT	TGTTCCATAG	TATATGCTGC	TCACTCATCA	3000
	TATTATTACT	ATAATGATCA	GCATGAATCA	CAACACCAAC	TTTACTATCA	CCTTTATGCT	3060
30	GCAAAACAGC	CTGACCAATA	TCAGAAGCGC	GGTCTAATAT	GACAATATCG	TCTCGGGTTA	3120
	AATTCAATCG	TTGTAAAAAG	TATGCAATAA	ATTCCGTTTT	GTTATACAAC	ACCGCATCTT	3180
	CAAACACATA	TATAGAGCTG	TCTCCATCAA	TATATTCGTT	ATAAGCGATG	GAACCATCTT	3240
35	GATTATAAAA	TTGTCGCATA	TATAATTTTC	CTTTATTATC	AGCTGGTGCA	TAATACTCAG	3300
	AAAATATGCG	CGTATAACTA	TAAAAATCTT	TACGTACTAA	CATACTATTA	ATTACAAATT	3360
40	CTGCACGATC	CACAATATCT	TTTTGTTTCAT	TTTGCAGATA	ACATGTTACA	AATGATGATT	3420
	TCCCATTAAA	ATATAGGCGG	ACTATCTTAC	CATTTCTTTC	TCTAAAACTA	ATGTCATGAC	3480
	CAAGCTCACG	TTCAATGTCA	TCTAACGTGT	ACGTTGTTGG	TGCTAAAGAA	ATATCACTAA	3540
45	AATACTGATA	CAACCAAATA	ACTTCTTGAT	CTTTAAACCC	AATGTTTTGC	GTTAATGTCT	3600
	GTATGTTCTC	TGACTGTATA	AAATCTAAAA	ACACAAATTT	AGTGTCTTGA	TTGTACGTC	3660
	TCAATAATTT	AGCACGGTAA	GCTTGTGCAT	ATTCAACACC	GCTACTCGCC	CAGCCTATAC	3720
50	CAAAGTTTAT	ATTATATATT	GTCATGCGCT	ACCCCTTTTC	ATTTATGGAA	AATGTATAAC	3780
	TGGCATACCC	TCTTTATCAA	ATGTAATCAT	GCTTTGACAA	ATATTTTTCa	CCATTCTTTT	3840
55	TTTGATATTT	CGTGTCAATA	CTTCAAATGA	ATCTAAGGCA	ACTCTATGGT	ATTCAAAAAT	3900

	GACTTGTTCCT AACCAACATG AATCAATTGC TTTCAAAAAG ACTTTTGTGAA CGAAAAATATT	4020
5	ATAATAATAT GCACTTTGCA TGTTTTTACG ATTCAAAGCT AATTGCTTTT CAAATTGCTC	4080
	TAATAAAAAAT GTCACTACTG CTTGCTTATC TTTAAAAATTA ACACAAGCCA CATCTTTATT	4140
	AAATTGGAAA CTTAATTTT GATAAATATA CTCGACAACA CGCGATTTTG TTAGCACCTT	4200
10	TTCCTCATTT ACAACATTT CAAATACATC TTTAGCTAAC GCTTTAAAAT CTTGATTCTC	4260
	AGCATCATCT ATTTCTAAAA CTCGATTGCG TTCCTCGTAT ACAAGATCTC GCTGTATACT	4320
	AATGCTTTTT TCAAATTCAT TAGCCATTTC ACGAGCTTTA ACCCCTTGTT CTTCCGAGAT	4380
15	aCGcTGCGCT TTAAC TACAA TTTGCTTAAC TTTGCGATTA AACAAATTAC TTTGCGATAA	4440
	TCGTTGTGCA TCTAATGAAT ATAATTGATT ATTTTCCGCT AAATTACTAT CGCTCCATCG	4500
20	CTTAAC TAAA TAATCATCTA GTGAAATATA TATACAAGAT GATCCCGGAT CCCCTTGTCT	4560
	ACCAGAACGA CCACGTAATT GCCTGTCTAC ACGGCTATTT TCCATATGTT CATGAATAAT	4620
	AACAGCTAAT CCACCTAATG CTTGACACC TTCACCAAGT TTAATATCTG TGCCTCGACC	4680
25	TGCCATACTA GTCGCAACAG TCATGGAACC AATTGCCCCT GCTTCAGCTA TCATCTGCGC	4740
	TTCTTTTGCA ACATTTTGCG CAATGAGTAA ATTATTAGGA ATATCCATTT GGAATAATAC	4800
	TTTCGAAAAG TATTCAGCCG CTTGAGCAGT TCTCGTTATG AGTAAAACCG GTCGCCCCGT	4860
30	TTCATGAAGT TCAACTATAT CATGAATCAT CGCGATGTTT TTCTCATCAA CTGAACGAAA	4920
	CACTTTATCT GGTTCATCGA TACGTTGAAT CGCTTTATCA GTTGGTACTT GTACGACTAT	4980
	TTTTGAATAC AAATCAAAGA ACTCTGATTC GCCTAATTTT CCTGTAGCTG TCATACCTGA	5040
35	AAATGATTCA AAAAGTTTAA ATAAATTCTG GAAGGTAATT GTTGCCATAA CACTTTTATC	5100
	TGTTGAAACC TCCATACCTT CTTTCGCTTC AATAGCTTGG TGAAGTCCAG CTTGCAACTT	5160
40	AGTTCCCGGT AACATACGAC CTGTAATACG GTCAATTAAA ACAATATCAC CATTATATAC	5220
	AAAGTAATCG ACATTAGATT CAAACAAATA TTGTGCGCGC AGTGCTAAAT TAATATTACG	5280
	CACTAGGACC ATCGCTTGTT CGCTATATAA ATCTTCAACA TTAAAGTATG ATTGTGCCGC	5340
45	TTCAATACCT TGATTTAACA GCCATATTTT TTTTTGGTC TTCTTCATTT TAAAATGCAC	5400
	GTCTTCAATC AATGTATCTA CAAACTCTTT CACAATATGA AATAGATTTG ATTGTAATCT	5460
	TGGTGCACCC GAAATAACTA ATGGTGTGTT AGCAGCATCT AAAATGATTG AATCCACTTC	5520
50	ATCAATAATA CCGTAATTTA ATTGTGGTAA AAATTTCCCT TCCGCACTAT CAGCCAAATT	5580
	ATCAATTAAA TAATCAAAC CGAGACGTCC ATTAGTTGTA TATATAATAT CATGTTCTA	5640
55	TATATTACGT TTTTCCCTT TTTGATACTC ATAATCCACA ATATCAACAA AACCTAATGA	5700

	TAATCATTCG TTGTAATTAA ATATGTTCCCT TTTCCCGAAA GAGCATTTAA ATATAAAGGC	5820
	ATCGTTGCCG TTAATGTTTT ACCTTCGCCT GTTTGCATCT CCGCAATGTT ACCTTCATGC	5880
5	AATACAATCG CTCCGATTAA CTGAACTTCT TTAGGATACA TACCTAATAC TCTCCAGCTC	5940
	GCTTCACGTG CCACTGCATA AGCTTCAGGT AACAAATGTAT CTAGTGTATC AACTCCTGAT	6000
	GCTAAACGTT CTTTAAATTC TATTGTCTTT TGTTTTAACG CATCATCAGA ATATGATTTA	6060
10	ACTTCATCGC TCCATGTATT GaTGSgTTcA CTATTTTTCT AATCGACTTT AGTCTTAATT	6120
	CGTTTTATCGT AACATCTAGT TTATGTTTCA TTTACTTCCC CACCATTGAG TTTCGATACA	6180
15	TCTAAGTAAT CTAAAAATCG TACTGGATTC ATTAAACGTG ACATATAATT TAGATGTTTG	6240
	TCTTGCTCTT CTTTAAAATA AACCTCGACA TTTGTATCTT TTAGTTCATG ATTTCTGGG	6300
	ACATGTTCTG TAAGCCATCC TTTTAAATCA TCATCTTCAT GGCTGTACG ATACACTTTG	6360
20	CAACCCAAAT GCTGAGCGAC ATAAGTTGCA AAAACATTG ACTTTGACCC ATAACATAATC	6420
	AAATTAATAG CCTTTAGGGT ATCTTGACTT TGCAAATCAT TCTTTAGTTG CTTAATATTT	6480
	CCCTCGATAT TGTCGTCCAT CCAACGTTCA ACGAGCCAAA CATGACCAAA CAGTTTCAAA	6540
25	AAATCATTCG AAATAGTTGG ATAGGTGTCA GATGGTCTG CAATAATGAC ATTGATCATA	6600
	TCATTTCCAT ATTGGTCATC GCCTATCTTC GTCACCCGCA TGCTTTTATA CTCTAAATCA	6660
	TATTGATGCG TCATCTCTGT GATTGTTAAA CATCTAAATA TAAGACTCGT CGATGCTGCA	6720
30	TTTCATCATTT TTATTTTATA AGCATAGGCK TCATCAGGAT ATTGAATCGT AATACTATTT	6780
	GACTTTACAA TCTCAGTACT TAGTTTTGTG CCATTTTTAT TATAAAAAAT GATGATAAAA	6840
35	TAACTGAAC CAGCAGGCGT TGCATCAAAA TCAAAATGCA ATTTATAATG CTGTCCTCTA	6900
	CGCAAAATTG GkAAACTTGG CGcACTTTTA TATTTTGAAA ATTGCTTTAA CATCAACCAC	6960
	TCATGAATCG GTAATCCAGA GGGCATCAAA GGATTTATAA AAGTCACTTC ACCATTTGAA	7020
40	AATGATACTT TAGAGCCATA CATAAATGTA GTTTGTGAAA TATAATTCCA AGTAACTTTA	7080
	AATGTTTTGT TTTTCAGCAT GTTGAACCTC CCCAAACTTG TCTTCCAAA TAATGTTGTA	7140
	AAAATTAACA AACCAACTTG CAATGGTAGG TGAATCATCA TTATGTCGCC CAGGAATACT	7200
45	GCGATTCATC ACTCTTGCTT GGTGTGCTGT CAATACAGGT AATAGCTCTT GAAATGCATG	7260
	TGGATCATAA TCATCATGTT GCATATATGC TATGGCAAAA ACAGTTTGTG ACAATGATTy	7320
50	CTTTTGAAAT GTTTGCCAAA ATTTTGTGAT TAATGCCTGT ATCGACGCTT GAGATGTATC	7380
	ACCTTCATTA GACACCAGGA CGTCTAATGC TGTACCGAAC TCTTCTGGTC TAAGTAATCG	7440
	CATATGTTCA GCAATCGTTC CAATATTAAAC AAGTGGTTTA CCAACAATAA TTGCCTGAGG	7500
55		

	TAATTCATGT GATTTAAAAT TCAGCTTTTC TAATGTCTCG TCAATAACAT TGATAATACC	7620
	TTGTTTCATAT TCAGATGAAC CGATATAAAA ACTACCACCT TCAACACGAG GATCGCCGAT	7680
5	AAGTAAAAAC GGTGCATTCA TACGTTTCAT CATATAATAT CCTTCGAAAC CTTCCGCTGT	7740
	TCGATAACCA CTAAAATATA CGTTTAGTGG CGGTTTCATA TCACCAGGGT GGAAATAATA	7800
	AATAAATTCC TGTCGTTGAC TATCTACGAA ACGACTACCA CCAAGTAAAA ATTGACCCAT	7860
10	GTCTAATCTA GACCATCGTT TGTGTATAGG TCCTAAATGT ACCGTCCCGT TCCCACGCGC	7920
	CTTAACAGTT ACACCTATAT AAGCATCAAA TGGTTTCGCA GGTATCTCTA AAGGACTGTC	7980
15	TAACATATCA TCAGTCAATA CGATTGTTC AATTAATGCA CCATCAGCGC CAGTCTGAAT	8040
	CAATCTAAAT GTATATTGCA ACTCGACCGC ACCATCAATA TCAAATTCTG GCCATATTTG	8100
	AATGACTTTA TCTTTATCGT AAACGAGATT ATTTTGCCAA GATGCGATAG GTTTAAATTC	8160
20	TTTCCCAAAT TCTCCACTCA ATGTGAGCTC TGAATTACCT TGGTAAACGA CATCTCCTTT	8220
	AAAATTCGGA TGCACAAGTG CTAACCTAGG AGAAACCTTA TCTCCATACT GTCCTGAGAA	8280
	GCTAACTGCC TCTAATTTAT TATTACGTTT TTCAATATTC CGGTAATGTA ATGGTTGAAC	8340
25	AACGTATTTT TGGACATTTT CGTCTGTGTC ATATTCAACT GACCAAATG ATTCATCAAC	8400
	ATACGTATTG TATGGTTCGC TTATCATTTG TAATAAATTC GTTAATGTCT CCGAGTATGG	8460
	TGCTTGAATA TAGATAAAAT CAAAGCGCCC TTCTGCTTCA ACAATCGCTT CAATAGCCTC	8520
30	TACATAACCA CTATCAAATT CAAACAATCC AATATCGAAG TAATCCCAAC TCACACCTTT	8580
	TTTGTGTGTA AAAATAGGTT CTAAATCGTC TCCTCCAATT TGCAAACTC TAAATTTACG	8640
35	TGGCATCATT TTCACCTTCT ATTAACTCAT CGAGCTGATT AATAATATTC TTAGAAGCAT	8700
	ATGCATCTAT TAATTTTAAA GAATAGGCGT ACGCATAATT CCAATTTTTC AAATAAAATA	8760
	AATAATAATT TAACGCATCA TCTAATTCAT CAACTGTATT TATAATACGG CCATTGTCAT	8820
40	AATCAGAGAC GTAATCTGTT TGTGACCAT TAATTTGTGG AATCCCAGCG CTAATTGCAC	8880
	TAATTTGTAA ATACAAGTCA GGTCTTTTGG ACATATCTAT CACAAGTCGC AACGTCCGCA	8940
	ATGCTTCTAC AACATCATGT TCAGCATGTA TCGTCTTAAC AGCAATGATG TCATCTTGAT	9000
45	CTTCAGGTGT CATTAATGCT GAAACATTAA CATCCGCATT CTGTTTAGCT TGGTATTCCT	9060
	CATTTACCGA CGTAATACAT TCACGAAGCC ACATCGGTAT GTCATTTTGA TGGCGCGATA	9120
50	ATAAAATTAA ACGGTAATAA TCTTCCTGTG CGATATAATC CACAAGTCGT TGCATCATTT	9180
	GTTGCAAATC AGCGTCACTC ATACCATCTA TCCATACACC TATAAATGTT TCCATCAATT	9240
	GACTACTTAT ATTAGGTGAT TGTCTCGTTT CAAATGGTGT GATTCGAATC ATTGTATTCT	9300
55		



	TTAAATGGGC ATTCTTTACG ATAGATTGAT ATTCCTCATC TGACACAGTT TCATTTCTAT	9420
	TTTTAAAAA TGAATAACTT AATGATTTTCG CTGGAATATG ATTGGCTATT TGTCGATTGT	9480
5	GCCTAGCATC TGAAGCCACA ATCACATGAT CATCTTCATG TATTTGTTGT GCAATCATTG	9540
	CTTGAAATTT TTCTTCAATT AGTTGAGCCA TATTGTTATA TTCTGTTTGT TGATAGTGAT	9600
	GTTGATATCT TTTTGAAACA GTGACTCTGC CATTTTTCAA ATCTTCATGA AGTACACAAT	9660
10	CTCCATTAAT CGTTAAATAT TCTTGGTAAG AAGCCTCTCC CTGATCATCA AAATAACGTA	9720
	TCGCTGATAA ATAACCTCTG TCATCAAAAA TATAACGCCG TTGTAAGTGA TCTCTTTCAA	9780
15	ATTCTTCAAA CCAAATTGAA TACCCTTCTT GACTAAATA AATATTTGTA TAGGTCTGTT	9840
	CACTCGTCAC ACATTTTAAT AAATACGGTG TGTACACAAA CTCAACATCA TCCGGCCATT	9900
	TTAAGTGATG ATAATTAATC GCTTGTGGCG CATGGTGAAT GAATCCTTGA ATTTTCATCAA	9960
20	ACACAGACGA ATACTTTGTC TCATATAAGT CATATCGATG TAAAAATGTT CTTAAATTTG	10020
	GTGCATGATT GAGAACAATC AGTTGATAAT CTAAGTCATT TTCAAGGTGC ATTCCCATTA	10080
	AACTAATCAT ATCGTCAAAT TCCGTCTTAT TTTGTAGTTG ATAATACGGC ACAGTCGTGT	10140
25	CTTGCCACCA TCGTTGGTCA TCGTACCAAG CTGGAATAAA GTATTTTATA ATTACCTCCT	10200
	TACCAATACT GGTTTAAAAA TGGCTTATAT TTATCAAAAT ATAAATATGT ACGAATTGTT	10260
	TCTGCAATAT TAATACTGAT GTAACTAAT ACAATCAGTT GTACTGAGAA ATAAATTTCA	10320
30	GTAGATAAAT GCGGTACAAA CAATGTGAAA TAAAGCGGTA TACCAATAAT GACTGTAACT	10380
	AATGCCAATC CAAACCAACA TACGCGTCGT GCTTGATAAT TTAAATAACG TTCTGTATCC	10440
35	TTACCAGGTT TAACTCCTGA AAAATAATTG CCACTCTTTA AGAAATCTTT GGATTTTTGT	10500
	TTAGTATTGA TTAAAAATCT CGATAAAAAA TAACCCAATA ACATTTGAAT CACTAAATAT	10560
	ACTGAAATAC CTACTGGACT ATCAAATGTC AGCATTGGCA TGTCATCTGA TATGCTTTTA	10620
40	TTAAACATAG ATAAATAAAA ATGAATGCCA CTTTTTAAGA AAACAAAAGC TGAAATACTC	10680
	ATCATTAAAG TAATACTGCC TGCAGGGTTA ACTTTCCAAG ATAAATAAGA TTTCATATTT	10740
	GTTGCGGAAA CGTTCATTAA ATCGATATAT GGTATTCTCA CTTCTACTAA TTCAATAAAT	10800
45	AATAAGATAA ACAATGTGAT TATCACAAGG ATGATTAACA ACGCAATCAC AATATGACTT	10860
	GCATCTATAT ATTCATTTTT TTGATGCATC ATTGATTTAA TAATACTAAC CATTACAATC	10920
50	GGCATTGGTC CTGCGATGCC GTAGCGACTA TTTTGTGTCAG CTAACCAAAC TAATAACATC	10980
	GTTCCAGTAA CCAAAATCAA TATTGTTAAG TAAATATTGT CTTGATGAAC ACGTTCTTTC	11040
	GAAACATATT CATGAATCAC AAAATAACTT TGAATAACAC TTAAATTTAA TGTTAAGATG	11100
55		

	GAAATCAGCA TCAAGATAAT CATTGATGTT AACCACGGAC CTAATCCTAA AGTGAAAATG	11220
	TTTAAAGTAT TAACGTCTCC ACCCATATTA GAAATAGCTA TTTTAAAAAA TGACTCATGT	11280
5	TTTACTTGCA TATCGTTaTA GGAAACGATG GAAATGTTTG TGCCTAATAT ATAAATAaAC	11340
	AAGATAAAAC ATGTGTATAG CATACGTTTA TATATAATTT TATATTCGTA TTGTTGTAAA	11400
	AGTTTAAACA TGTTGCACCT CTTTATATC AAAACATTA AAAAGACTAA GGGTTCATCA	11460
10	CTAATTATTA AAATCCTATA TCGATTTTTC TAGTGATTGG TGCCTCAGTC TTTTAAATTT	11520
	TAGCCAGCTA TAAATTCAAT TTATGCTTGA GAATCATCTT GATCATTTTC ATCTTTCTTT	11580
15	TTCTTTCTCT TCATTAAACC TAAACCAACT AATAATGTCA TAACGCCACC TAGTAATCCA	11640
	TTTTGTTTTA TTGAGTCACC TGTATCTGGC AATCTTTTTT CACTTTGTGC TGGTGTGCCA	11700
	TTATGTTTAG TCACTTCAGA TGTTGCACTT AATGTAGACT GAGATTCACT CGTGCTCGTT	11760
20	GTTGCTTCAC TTGATAAGCG AGATGTGCTC GTGCTGTGAG TATGATGCAT ACTCATTGAG	11820
	TCTGACGGAT GCATTGAGTT AGATTCAGAT GTACTTGTG AGCCGGACAT ACTTGTGAT	11880
	GTTGAGTCAG AAATGCTTTG TGAACCAGAC ATAGATGTAC TCAGTGATTC GGATGTGCTT	11940
25	GTCGAATCGG ATGTGCTCAA TGACGTTGAT GTGCTTGTG ACACTGATTC TGAGTCACTA	12000
	ATTGATGTTG AGTCGGATTT GTCTTGTGAC ATTGAAACAC TCGATGAATT AGATTCATC	12060
	ATTGATGTTG AGTCAGATAC GCTCGTTGAA CCTGAACCAG ACGTACTTAA TGATTCAGAT	12120
30	ATGCTTGTTG AAGTTGAACC ACTTGTGAG TCCGATGTAC TTGTCGATGT CGAGTCTGAA	12180
	TCTGATGTAC TCAATGATTC TGAGTCACTG ATAGAAGTTG AATCACTTGT AGATTCTGAT	12240
35	TCTACTGTAC TTTGTGAACC ACTGATACTT ATTGAAGTAG AATCACTGAT ACTGTCTGAT	12300
	GTTGATAATG ATGTCGACAC CGATGTGCTT TGTGATGACG ATGTACTAGC ACTCATTGAC	12360
	ATTGATGTTG ATATCGATGT ACTTAAGGAA CCAGATGCAC TTGTACTTGT TGA CTGGCTT	12420
40	TGTGACATTG AATCACTTAA TGATGTAGAT GTGCTTGTG AGCTCGAGTC ACTTACACTT	12480
	GTTGAACCTG ATATTGAGTC ACTTAAACTT GTCGATGTTG AACTGatwC GcTTCCGCTC	12540
	ATTGAGTCAG ATGTTGAAAG TGATGTACTC GTTGAATTTG ATCCACTGAT GCTAGACGAA	12600
45	TCACTGTAG ACATTGAGTC GCTTTCTGAT GCACTGATGC TCATAGAGTC AAATTGACTA	12660
	TTACTTGTTG AGCTTGACTG CGAATCGCTC ACACTTGTTG ACGTTGATTC TGATCCACTC	12720
	ATACTTTGCG AGCTACTCAA TGATTTTGAA TCACTTAATG AATCCGAAGT GCTAAGACTT	12780
50	GTGGAACCAC TTAAAGATAT TGATCCACTT AATGAGTCGG AGTCACTTGT ACTAGTAGAA	12840
	TCACTCATTG ATATTGAATC ACTTAGcGAG GTAGACTyGc tTACGCTTTC TGAACCACTT	12900

55

TTTGAATCAC TTAATGAATC AGATTCACTC ACGCTTCTG AACTTCTTAG TGACGTCGAT 13020  
 ACACTTAATG ATGACGAATC GCTTGTGCTT ACTGAATCG 13059

5 (2) INFORMATION FOR SEQ ID NO: 221:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10758 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

AGGGATGGCC TTACCTAAAA AACCGGGnAA ACCCTCCAAA ACCCATTTAAA AGGnTGGnTA 60  
 CCCTTTAAAA TGGTAGCATT TAACCGCCAC CCGCCAAGGT GGGTGGTTTA TTCTTCCGTT 120  
 20 ATTTAAATTA GTACACCATG CAGATTCTGT AGTTGAGGGA TATTTTAACG AAAGCTTATT 180  
 AGCAACTGAT AAAAAAATAC GTCCTAAGGC ATATATTGCT TCATGGAAGG ACATCGAGCC 240  
 GGCTAAGAAA ATAGAATTTA AAATTAAAAA AGGTATTAAA TGGCATGATG GTAATGAATT 300  
 25 GAAAATTGAT GATTGGATT ATTCAATTGA AGTCTTAGCT AACAAAGGACT ACGAAGGTGC 360  
 TTATTATCCA AGTGTAGAAA ATATCCAAGG TGCGAAGAT TATCATGAAG GAAAACTGA 420  
 TCATATTAGC GGATTGAAGA AAATAGATGA CTACACTATG CAGGTTACAT TTGATAAAAA 480  
 30 ACAAGAAAAT TACTTAACAG GATTTATTAC TGGACCTTAA TTAAGTAAAA AATATTTATC 540  
 AGATGTACCA ATTAAAGATT TAGCGAAATC AGATAAAATC CGAAAATATC CTATTGGTAT 600  
 35 TGGACCGTAT AAAGTTAAGA AAATCGTTCC AGGTGAGGCT GTTCAACTCG TTAAATTTGA 660  
 TGATTATTGG CAAGGTAAGC CTGCACTAGA CAAAATCAAT TTAAAAGTTA TTGATCAAGC 720  
 GCAAATTATT AAGGCAATGG AAAAAGGCGA TATTGATGTT GCGAATGATG CTACCGGTGC 780  
 40 AATGGCAAAA GATGCTAAGT CATCTAATGC TGGTCTCAAG GTATTATCTG CGCCAAGCTT 840  
 AGACTACGGT TTAATAGGtT CGTATCTCAT GATTACGATA AAAAAGCTAA TAAACTGGT 900  
 AAAGTGAGAC CAAAATATGA AGACAAAGAA TTACGTAAAG CAATGCTTTA TGCAATTGAT 960  
 45 AGAGAAaAAT GGATCAAAGC GTTTTTCAAT GGTACGCTA GTGAAATCaA TAGTTTGTGA 1020  
 CCATCTATGC ATTGGATAGC AGCCAATCCT AAGGACCTAA ATGATTACAA ATATGATCCT 1080  
 GAAAAAGCTA AAAAAATCTT AGATAAGTTA GGTATAAAG ATAGAGATGG TGACGGATTT 1140  
 50 AGAGAAGATC CTAAAGGTAA TAAATTTGAG ATTAACTTTA AACATAATTC AGGTTCTAAT 1200  
 CCTACTTTTG AACCAAGAAC TGCTGCGATA AAAGATTCTT GGGAAAAAGT TGGCTTGAAA 1260

55

	AATACGATTC CTGTTTATAT GCCATATATC ACATCTTATT TCATGACGCG TGCTATCGGC	1380
	GACAGACCTT TAGTCGTCCC GCATCAATCT CAGAACTTAG CATTTATTGG TAACTTTGCA	1440
5	GAAACAGAGC GAGACACTGT ATTTACAACA GAATATTCGG TTCGTA CTGC CATGGAAGCT	1500
	GTTTATCAAT TACTAAATAT AGATCGTGGT ATTCCAGAAG TCATCAATAG TCCATTTGAT	1560
	CTTCGCGTCT TAATGGATGC CATATACGAA CTGAATGACC ACCAAGATT GCGTGAGATT	1620
10	ACTAAAGATT CGAAAATGCA AAAACTCGCA TTAGCAGGAT TCCTTAAAAA GATAAAAGGT	1680
	ACGTACATTG AGTCATTATT AAAAGAACAC AAATTGTTAT AACGAAAACC ATTAATAGAT	1740
15	TTTTATTTGG TGATTTCAAA TCATGAGACT GGGACAGAAA TGATGTTTTT ATAAAAATTA	1800
	TTTCGTTGTT CCACTCTCAT GATTTTTTTG ATGAAACATA ATTACATGAT TGATTGCATC	1860
	ATTTTGTTAA ACAAGTGATT GCAAACCTGC CATTTACAC TGAAAATTTA CATAATAAGT	1920
20	GACGATATTT TACAAGTCAT ATACAAATAA CATATATTGT TAAATAATTT TACCTAATCT	1980
	TAACATTAAA TTTACAATTA TAAGCGATAA TCTAAATATA AAGCTTATTT GAGGTGAAAT	2040
	AATGGAAATG TCGGTTACAG AAGTCATTTT CTCCTTTTTA GGTGGTTTAG GTATTTTCCT	2100
25	TTACGGCTTA AAAATCATGG GAGACGGGCT TCAAGCATCA GCAGGAGACA GGCTACGAGA	2160
	TATTTTAAAC AAATTTACAT CAAATCCAGT ATTAGGTGTT ATTGCAGGTA TCGTTGTAAC	2220
	TATTTTAATA CAAAGTAGTT CAGGTACGAC AGTTATCACA ATCGGACTGG TAACAGCTGG	2280
30	ATTTATGACA TTGAAACAAG CCATTGGAGT GATAATGGGT GCTAATATCG GAACAACGGT	2340
	AACTGCATTT ATTATCGGTA TAGATTTAGG CGAATATGCA ATGCCAATTT TAGCATTAGG	2400
35	TGCATTCTTA ATCTTTTTCT TTAAACGCTC TAAAATCAAT AACATTGGCC GCATACTATT	2460
	CGGTTTCGGT TCACTATTCT TCGGTCTAGA ATTTATGGGT GATGCCGTTA AACCTTTAGC	2520
	ATCAATTAGAT GGATTTAAGC AATTAATGCT TGATATGTCT ACAAATCCAA TACTCGCTGT	2580
40	CATTGTCGGC GCAGGGTTAA CAGCACTAGT TCAAAGTTCA AGTGCGACGA TTGGTATTTT	2640
	ACAAGAATTT TATCAACAAG ATTTAATTAG CTTAAACGCA GCAATCCCTG TGTTACTAGG	2700
	CGATAACATT GGTACCACGA TTACAGCTAT CTTAGCTAGT TTAGCCGGCT CAATCGCTGC	2760
45	AAAACGTGCG GCGCTTGAC ACGTCATCTT TAACTTAATC GGGGTAATTA TCTTCACAAT	2820
	TTTCTTGCCA GTTGTGATT CATTGATTAG TTTGTTACAA GATTTATGGC ACTTAAACC	2880
50	AGCGATGACG ATTGCAGTAT CACATGGTAT CTTCAACATA ACAAATACTT TGATTCAATT	2940
	ACCATTTGTA GCAGGTTTAG CATGGATTGT TACAAAGCTT GTCCAGGTA AAGATATTGC	3000
55	TGATGACTAT AAACCTCAGC ACTTAAACAA AGATCTTGTT TATCACGCAC CTGGTGTTC	3060

	AGACATTTCGC	GAAATTACAA	AAGACGATAA	AAAATTGATC	AAAAAGCTTG	AACAAAAGCA	3180
	TCAAGCTGTT	GAAACAATCA	ATGATAGCAT	TCGAAATTAT	TTAGTTAGAA	TTTCTACAAA	3240
5	AGCCATTACG	AAGGCAGACG	TTGAGCGTTT	AGCAGTTATG	TTTGATGTCA	ATCGCTCTAT	3300
	TTTAAAAGTA	GCAGAGCTAA	CAGAAGAGTA	TGTCGCTCAA	TTAAAACGCC	AACATGATGA	3360
	AGATATTTCGC	ATTACAGAAG	ATGCACAACG	CGGTATGGAT	AAATTATTCA	ACCATGTTGC	3420
10	TGAGTCATTT	GATAAAGCCA	TCGACATGTT	AGATGTTTAT	GACAAAACGA	AAAAAGATGA	3480
	AATTGTAGAA	CGTAGTAGAG	AATCATTTAA	TATTGAACAT	AACTACGCA	AAGGTCATAT	3540
15	TAAACGCCTT	AATCGTGGTG	AATGTACAAC	AAAAGGCGGA	TTACTATATA	TCGATATGAT	3600
	TGGTGTTCCT	GAACGTATCG	GTTATCATTC	ACGAAATGTT	TCTGAAGCAC	TTGTTGGCCT	3660
	TAACGATGAT	GTACCTACAG	ATGAAGAAAT	TGCAACAACT	GAAATTTAAT	TTTTACTGTC	3720
20	TTATTTATAT	TCATATTTTT	TTAAAATTAG	AGATTCAGAT	GCATGTAAAA	AGCCAATCCA	3780
	ACATTCATGG	GTTGGCTTTT	TTGTTTAGCA	AAATTTATTA	TCTTAAATCG	GCTATAAACA	3840
	CTGATATAAT	AATGCTTCAT	TAGTATGCGG	TAAGCATGAC	GGACACTGTT	CTCGGAGTCT	3900
25	GACCCCGAAA	CGTTTAATAT	ACACTTTTAC	ACGTCGCCTT	CATTGAAGCG	AATTGCCATA	3960
	ACCTTCACAT	TATATATAGT	TCTTTCCATA	TAAATGTCCA	AATTTTTAGA	ACAACGCAAT	4020
	AAATAACCAT	CCACCTAACT	TATCAAAAAT	TTAAGTGGAT	GGTTTTTCAT	TTTCATTTAT	4080
30	ATTTATATTA	GTGTTAATCC	AATCATAGAT	TTATCTATAT	GCACTGCTCT	ATACATTTCC	4140
	TCATTTAATT	TGCTTTACTT	TCATTTATAT	CATTATCAAA	ACACTTGGCG	TGTCATCGTT	4200
35	ATTATTTTCGC	ATCTTTGACA	CGTTTATCAT	CATTAGGAAT	CGCGAATAAA	ATTGCGATAA	4260
	ATGCCATGAT	TCCCATTAAT	ACGTTAACCC	AAAGTGCAAT	CATCGCACCT	GTATGAATGC	4320
	TCGTTGCAGC	AACTGCACCA	GCATATACAG	CACCACTAAT	TGCGACACCG	AATGCGCCAC	4380
40	CAAGTGATGA	AGCCATTTTA	TAAATACCTG	AAGCAACGCC	AACTTTATCT	AACGGTGCAT	4440
	TCGAAATAGC	TGTATCTGTA	GAAGGTGTTG	CATAAATACC	TAAGCCTAGT	CCGAAACATA	4500
	AATATCCTAC	GACACAACCTG	ATAACATAAA	ATATGCCTGG	TAAGAATACT	AATGAAATAA	4560
45	GTGCAATACC	AATGACCACA	ATGAATGTAC	CTAATAACAT	TGGTCGCTTA	GAACCCATTT	4620
	TTTGTAATAA	TTTTTCACCA	ACTCGAATCA	TCAATAACAC	CATGATTAAA	TAAGTAATTG	4680
	ATAAGTATCC	TGCCTGCAAT	GCTGTATAAC	CTAAACCTTG	TTGCACGAAT	GTATTCGCTA	4740
50	CAATTAATGT	ACCTGCAAAA	CCGTTTAAATA	AGAAGTTCGA	AATCGTTGCA	CCTGTATATG	4800
	GTTTATTTTC	AAATAATTTA	AAATCAATAA	GTGGATTATC	TACTTTTTTC	TCAACATTTA	4860
55							

	AACCAAGTGC	TGCACCTTTA	GTAATGACAA	CGTTTAACT	TAGCAACATA	ACTACTAGAA	4980
	CAATTAGCCC	TGCAACGTCA	AATTTATGTG	TATTGGTAAT	TTCTGATTC	GTTTCAGGCG	5040
5	TCCCTTTGAT	GAGTAACATT	GAAAGTACGG	CAACGATAAT	TGAGAAGATG	AAAATCCATC	5100
	TCCAACCCAT	AGTTGTCGCA	ACTGCACCAC	CGAAGAGTGA	ACAGATACCA	CTGCCACCCC	5160
	AAGAACCGAT	AGACCAATAA	CTTAAGGCAC	GCTGACGTTT	AGCACCCCTGA	TAATAAGTTT	5220
10	TCATAATGGC	CAATGTAGAA	GGCATAATAC	ACGCTGCTGA	TACACCTTGT	ATAACACGAC	5280
	CTAAAAATTAA	TAATGCCGGT	AAATTCGTAA	TAATAATTAA	TGCTGAACCA	ATAATACTTA	5340
15	ATAATAAACC	GATATTCGTC	ATTTTCACGC	GCCCAATTTT	ATCTGCCAGA	CCACCTGCTC	5400
	CAACAACAAA	CATGCCTGAA	AATAGTGCAG	TTAGACTGAC	CGCAATACTA	ATTGTCCCCA	5460
	TGTCTGTACC	AAAACCTTGT	TGTAAATTCG	GTACAACATT	TACAAGTGAT	TGTGCAAACA	5520
20	ACCAAAATGT	AATAACACCT	AATACAATAC	CTAAGATTAA	CTTGTTGCCC	CCGCGATACG	5580
	TTTCATTCAT	GTTAGTTATC	TCCTTTAAGG	TAATCTAAAA	CAACTGTCCC	TACTGCTTCT	5640
	GCAGAAATAA	GTAATGATTT	TTCTGAAATG	TTAAATTTAG	GATGATGATG	TGGGTAAATT	5700
25	TCACCATTTT	CCACCGCTGC	ACCTGTATAA	ATAAAGGCAC	TTGGGCGTTC	TTTAGCATAA	5760
	TATGCAAAGT	CTTCTGAAGG	TGGTTGTGGT	TCACACATTT	CAACACCAAA	ATCAAGGTTT	5820
	GCTTCTTTCA	ACGTCTTAGC	CACGTACTCA	GTAAACTCTG	GATCATTATA	TAATGCTGGA	5880
30	TAATCATCGT	TATATTCTAA	GGTGCAAGT	ACACCATACA	TATCCTCTAA	TCCTTTTGAT	5940
	AAACGTTTAA	TTTCTTTTTC	AATTGTTGCT	TTTGTAGCAT	CTGTTAATCC	ACGTACATCA	6000
35	CCTTCAATTT	CAACAACATC	TTTAATGACA	TTGAATTGAC	CTTTACCGTC	AAATGAACCG	6060
	ATTGTGACAA	CACCGGTTTC	AAATGGACTT	AGTCGTCTAG	ATACAACTGT	TTGTAACGCT	6120
	GTGACGAAGT	AGCTACCTGC	AACAATGGCA	TCATTGGCCA	TATGTGGTGA	TGAACCATGA	6180
40	CCACCTTTAC	CTTGAACTTT	CAATTTGAAG	AATGCGCGTC	CTGTTTGAAC	ATAACCAGGT	6240
	CTGTAATACA	CTTTACCTGT	TTTCATTGTG	CTCATGACGT	GTACACCTAA	TACATGATCA	6300
	ACACCGTCTA	ATACACCATT	TTCAATCATT	GTTTTAGCAC	CACCTGGTGG	TACTTCTTCA	6360
45	GCTGGTTGAT	GTATCACAAC	GACTTTTCCT	GTAAACTAT	CTTTCATTTT	AGCAAGCGTC	6420
	TCTGCTAATA	CAAGCATGTA	TGCTGTATGT	GCATCGTGAC	CACATGCGTG	CATAACACCT	6480
50	TTATTTTGTG	ATGCAAAAGA	TAATCCTGTA	TCTTCAGTAA	TGGGTAATGC	GTCAAAGTCT	6540
	GCACGGATTG	CTAATGTTTT	ACCAGGTTTC	CCTGAATCAA	TCGTTACTTT	AATTCCACGT	6600
	GGTCCGACAT	TCGTTTCTAC	TTCCACATCT	TTACCTTTGT	AAAATTCAGC	GATGTATTTT	6660
55							

	ATCATTTTGC CTTCTTTAGA TTTTAAAGTT TCAATTAATT GTTGATTCAT ATCCTTCATC	6780
	TCCTTAGTTA CATCATAAAT GATTAATCAT TATTTATATT GCCAACAACA GAGATGTTAA	6840
5	CCATTAATTT TTTGCAATTT TAGCTTTGAA TATAAAAAAT CACAAATTAT GTATATCAAA	6900
	ATTTGTGATT TGTGATCATT TTATGAACTT GGGTAACGTT TTACTIONCAAT TAAGTGAATC	6960
	CCATTTCGTAA TCATTTTAAAT GTTTAATGCC AGTGTGTCCG TGATATCTAT ATCATATACT	7020
10	TCTAATTTG GAAAACTCAT TCGATTAAACG TAATCTATAG AGTCCTTGTC CATGCCATGT	7080
	ATCGTATGAT GTTTGCGCCA AAGATTAAAT AACGCACCAT TTTCTTTATC TAAGGTAAAA	7140
15	TGTTTAACTCT TATACATACC TTCTTCCAGG GCATTAATGT TCAAATGAAT CATTTCGGTC	7200
	GCACGCATAT TCATTTGATT GTCCAACGCT AAGTACGGAT TAAAATGCTT TGCATCATAT	7260
	AACAATATTT GAAAATTTGA ATCAGTCCCC GTGACAATAC ATGTATCATC AGAATACAAA	7320
20	ATATTGCTTG TTAATTTATT AAATAGCAAT GCCGTGAAAT AGACCGGACG TTTTCCATTA	7380
	TATTGATGAA ATAGTTCAAT AGAATTCATA TAATCCCGTT CATTTTTACA ATGACTGACG	7440
	TGCAAATCAT AATTCAACCA ATACCCGATA CCCTCTACTT TAGAACTTAA TTTTAATAAT	7500
25	TGCTCAATGA TGATACCACC TCTAAAATAT TCGCCGTTTG TAATAAATGT ATCACCCGTC	7560
	AATGTATTCC AATTGAGTAA AATGAGTGGA CGCTTTAGGC GATGACGATG CATTAAAGTCG	7620
	ATAAGGTAAT TCGTTTTATT AATAATCATT TGACTIONCGCG TTTTAAATTC ATCATCATTC	7680
30	ATTTTATTAA AATCAACAGC GTCATTTGAA TTGGCATCAA ATACAAAATG GTCGATGTGT	7740
	GGCTCAAGTC GTTTCAATAA TGGTAGATGT CTTTCCGTAG CTTGATCTAA GTGAATGTAC	7800
35	AAGCCACCAT TAGGGAATAA TGCTTTAAAA TAATCAATCA TTTCAATCAA AGACGTGTGC	7860
	AATGTCGTCA CATACAAGTT GAACTTCAAA TCTTTTCTAT GACTGACATG CAGGGCAACG	7920
	TGATGGATAA AAATTTTAAA TGCATCGATA TAATCACGTG AGTCATACTG ATCCAAATGC	7980
40	ATGGTCAAAC TAAAGTTATG ATCTAATAAA AAGTCTAAAC ACAAATCAAT ATCATAAAAT	8040
	ATATTGAAA TTTCTGCATC ATACGTGAAT GGCGCATTGA GCTTTTTTCAT GATATATGGA	8100
	ATCACATCAT ATGCTAATAC TTCATTGACT TGAAAATCAT GATGACATGT AAGCAACTGT	8160
45	GATTGATACT GTGTATTGAG CAAATTCCTC AAATAGCCCA CTTGAATAAT ATGATTAAAT	8220
	TGATTTAGTT GGTGATTGGT TGGTTGAAAG GCAATCTCTT TATAGTTCAT CTTTTCAATA	8280
50	TCTTCAATAA AATGATTCAT TTCTTCAATG TAGTCATTTA AAAGTAATAT CAATTCACGG	8340
	TCGTGATAAT CATGTTGTGC CGATTGCTGG TTTTCAGTGA TTGCTGGACG ATCACCTCGA	8400
	TATTGTTTAG GTGTTTGATG CGTAAAGTGT TTAAATGTTT TCGCAAAGCT CGCTGCACCT	8460
55		

	TTCGCATGCT	CAATTCGCGT	CGTATTTAAG	AAATGATGGA	ATCCTACACC	TAGCGATTCT	8580
	GTAAACTTTT	TAGACAGATG	GCTCTCTGAC	CACCCAACGT	ATTGCTTAA	TTCTGAAAGG	8640
5	CTTAAATCTT	CATGAAAATG	TAACTCGATA	TAGTCGCATA	CTTGATTAC	TTTATCATCA	8700
	TTTAAGATAC	TTTGGTTCGA	ATGATATGTA	CGCGGGACAT	AATGAATCAT	ATGCATAAGC	8760
	AACTGAATCA	CAAGTTGTTG	CTCAGTCAAT	TTAGACAAC	CATTATGTCG	GATATGTGTT	8820
10	GAAACCAGTC	TTGCCATTAT	ATTTCTCAGT	TGATGTATAT	TCTTTGTTGT	GGTCGCATCT	8880
	GTTAAGTGAA	AATATAGACA	ATGCACATCA	TCAAACCTGT	CTGCTAAATA	TTTCATTTGG	8940
15	AATTGGATAT	AACATATGAT	GCCATCTTGT	TGAAGTTGAA	ATCGATACAA	GTCGCGGTGG	9000
	TTAATGATGA	AAATGTCGCC	ACTGTTGCAT	TGCGTCATAT	TATTTTTCATC	ATAAATGTGT	9060
	GCCTCnCTT	TAATAACAAA	ACCAATCATT	AAACTATTGA	GCCTTTTGAA	ATCTGACATA	9120
20	CTCTCAGTTT	CTACTCGAAT	TAAATAATCA	CGTTGCATAC	TATCCCTCAA	TTCAGTAATA	9180
	TGAATACGTT	TATTTTACAT	TATTTTACAG	CAACATATTT	GAATTTTATA	TTGAATCGTG	9240
	TGTGTGGATG	ATTATTTATC	CTCACTCGGT	TCAAGATGTA	GACTATCAGT	AAAAAAGTA	9300
25	TTTTACCTT	TTTTCTCCAC	AAAAGTAAAT	TCAATGTCTT	TATATCCAAC	TGrTGaACCT	9360
	TTTAAGTCTC	CCgAACCTTT	CaACaATAAC	TTTGGTGCTT	TATTCGTTGG	TATTTTATAT	9420
30	CTTTTTCGTA	ATTGTTTTAC	ATTATAGTCA	TCATTAGTTA	ATTGATATTT	TGCTGAATAA	9480
	CTCGGTACCT	CTGGATTATA	TGATATATCG	CCGTCTTTGT	ACTTCGACAA	ATCTTTAAAG	9540
	CTGCCATATT	GCgCGAAGAA	CTTAAATTC	TCGATTTCTT	TTTTTATATT	TTCGTCTTTG	9600
35	ATACCTTTAG	TTGGAATGAT	TTTATTGTCT	ACCATTTTAA	CGGGATATTC	TTTATCTTTA	9660
	CTCTTAGGTC	TACCATCTTC	ATCATGAAGT	GTTTCACTCA	CTATATACTT	CCCGGTTGTA	9720
	GTCTTAGTGT	TTCTATTCT	ATATAGAACC	ATACCTTTTG	ATTTCATACG	TTCCCTTTTA	9780
40	GGTTGAACAA	CCATTTTCTG	ACCAATAATC	CATGTACCTT	TATCATTTT	ATCAAATTCG	9840
	TCATCAGGAT	AACCTTCTTT	ATCGTATAAA	TCCTCTAGAT	TTTAAATCGG	ATACATACTC	9900
	AATGTTTTTT	CAAAGCTTTT	CTTAACTTCC	GCTTCTTTAC	CTATGCCACA	ACCAGCAGTG	9960
45	AAACTAATGA	CTAATATCAA	AAACTAATA	TACAATACCA	ATTTGTTTAA	TCGTTTCATA	10020
	ATTTACAAAT	CCTATTCTTC	TTATTATCTT	TCCTGGATTG	ATTTCATATT	TTGATCGAGT	10080
50	CATGATTATT	TATCCTCACT	TGGTTTAAAA	ATTAACCCAT	CACTAAAGTA	AATGTTCTCT	10140
	TCTTTTTTCT	CTACAAACGT	AAATTCAATG	TCTTTATATC	CAACTGATGA	ACCTTTTAAA	10200
	TTCCCTGTAC	CyTTCAACAA	CArCTTCGGy	GCTTTATTTG	TTGGTATGTC	ATATCTTTTA	10260
55							



ACCTCTGGAT TATATGATAT ATCTCCATCT TTATAATTCA TTAAATCTTT AAAATTGCTA 10380  
 TATTGCGCAA AAAACTTAAA GTTTTCGATT TCTTTTTTTA TGTtTTCTTC TTAACTTCC 10440  
 5 TCAGTAGAAA TGAATTATT ATTAAATCATT TTAAGTGGAT ATTTTMTTTG ATTATCCTGA 10500  
 GCTACTTCGT ATTTCTCCGT CTTTAtTTCA TTAGTATAGT AAAAtCCTTT TGCACTTCTT 10560  
 GTATTTCTAT CTATCTTCAA AAGCATGCCT TTTATTTTTA GAGCTTCTCC TTTATTTTGA 10620  
 10 ATTGCCATTT GAGAATTAC AATCCATGTT CCCTTATCAT TTTTATCAAA TTGATCATCA 10680  
 CGATATCCTT CTTTATCGTA TAAATCCTCT AGATTTTTAA TCGGATACAT ACTCAATGTT 10740  
 15 TTTTCAAAAC TTTTCTTT 10758

(2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1109 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

nTaTCaACTT TGGaATTtAA AgTCAATAAC TTTTTTAAAA ACTTTTGTG TTCACAACCC 60  
 30 GCTTCTTTTT CAACGCGTTT ATTGCTTAAC ACAAGAACTT ATTTTACCAG CATTCCAAAA 120  
 CAAATCAACA TAAAAACGTA CAAAATAAAA GTAATTTTGT ACGTTTAGCA TATATTATAC 180  
 CTATTTATTT GTAGCAGCTA TAACTTTTTG TGCAATCGAG CTATAAATTT TACCTAGACG 240  
 35 ATCATCTGAT TGATATATTG ACGGTGCAAA ATCTTTTGGG TTCCAAGATG GTTGCTCTAA 300  
 AGGTAATTCC CCAAGTAATT GAGTATTAAG TTCATCAGCT AACTTAGTAC CGCCACCTTT 360  
 GCCAAAGACA TATTCTTTAT TACCCGCTCT TTTACTTTCA AAATAACTCA TGTTTTCAAT 420  
 40 TACGCCAAGA ATAGAATGAT CCGTATGTTT TGCCATCGCA CCTGCGCGAg CTGCAACAAA 480  
 TGCTGCTGTA GGATGAGGTG TCGTTACAAT AATTTCCCTTA CTTGAAGGTA ACATCGTATG 540  
 AACATCTAAA GCTACATCTC CTGTTCCAGG TGGAAGATCG AGTATTAAAT ATTCAATGTC 600  
 45 TCCCCATTTA ACTTCTGTAA AGAAATTCGT CAACATTTTA CCTAACATTG GCCCTCTCCA 660  
 TATAACTGGC GCATTTTCTT CCACAAAAAA GGCCATTGAT ATAACTTTAA CGCCATGACG 720  
 50 TTCAACTGGA ATTACTTCCT TCCCTTTAAT TCCAGGCTTT TCATCAATAC CCATCATATC 780  
 TGGTACACTA AATCCATATA TATCGGCATC TACTAATCCG ACTTTTTTCC CTTACAGAGC 840  
 TAAGGCAACG GCTAAATTTA CTGCAACAGT AGATTTACCG ACACCACCTT TACCGGAGGC 900

55

ATTTTCTTCT TTTGGTTTAA ATTGATTAC TTTTCTTCC GGCAATGTTT CAAATCGTAT 1020  
 ACCGACCGTT TTCGCACCGT TTTCTTTTAA TGCATTAACA ACAGCCATCT GTAAATCTAA 1080  
 5 aTTGCGtGCA CCACCTAATT GTGCCATTG 1109

(2) INFORMATION FOR SEQ ID NO: 223:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3997 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

TCTTTATTTA AAAAAATGAT TGTCTAGTTT GTATCTCTCT GAAGATTTGG CAATAAATAA 60  
 20 AAGCCGATAA CCGTATAATG ATTATCGACT TAAAGTTTAT GTGGCATTTT TTTACTTTTGT 120  
 AATTTCAAGT GAGTTAGATG ATTATTATCA GATAGATTAT TGCTTATAAT CATATGATGT 180  
 TTGAATGATA TCTTTGATTT CACTGATTAG TGCTTCTTTA GGATTAGCAG TTGTACATTG 240  
 25 ATCTTCAAAT GCGAGCTCTG CCATTCTATC AATTGACTCA TTTAATTCTT CTCAGACAC 300  
 ACCTTGATGAT TTCAAATTCA TTTCAATTCC GACTGATTGA CCTAATTCGT AGACAGCTTT 360  
 AgCTAATGAT TCTACGAGTG CTTCTGTCGT ATTACCTTTT AATCCTAAGA ATTTGGCAAT 420  
 30 ATCTGCATAA TCTGTATCTG CTCTGAAGAA CTCATATTTA GGGAATAATG CATGTTTTTG 480  
 CGGGTCTTTG GCATTATAAC GGATAATATG CGGTAGTAAT ATCGCATTCT CTCTACCATG 540  
 CGGAATACCA TATTCGCCAC CAATTTTATG CGCAATTGAG TGTGCAATGC CTAAGAATGC 600  
 35 ATTTGCAAAT GCCATACCAG CCAAAGTTGA TGCCTTATGC ATTTTCTCTC TTGAAACTTT 660  
 ATCAcCCTTT TCAACAGATG ATTTTAAATA TTCGAACGTC AATTTAATCG CTTGTAGACT 720  
 40 CAAAcCTCTT GTGTAGTCTG AAGCCATTAC AGATACATAT GATTCCATTG CATGCGTTAG 780  
 TACATCCATT CCTGTATCTG CTGTAACGCT TTTTGGCACA CTCATCACAA ATTGAGGGTC 840  
 AATAATTGCA ACGTCAGGTG TTAAAGCAAA ATCAGCCAAC GGATATTTTA CATTTGTTTC 900  
 45 ACTATCTGTG ATAAGTCAA ATGGTGTTAC TTCTGAACCT GTACCTGATG TCGTAGGGAT 960  
 ACAAAATGAAC GTCGCATTTT CAGGCATGCC TATTTTATAA GTACGTTTAC CGATGTCTAG 1020  
 GAACTTTTGT TTAGCACCGA AGAATGATGT CTCAGGGTGT TCAAAGAACA TCCACATTGC 1080  
 50 TTTTGCAGCA TCCATCGCTG AACCACCACC AAGTGCAATG ATTGTATCCG GTTGGAATC 1140  
 AACCATCATT TCCAGACCTT TATATACTGT ATTAGTTGAT GGGTTCGGTT CGACTTCGCT 1200

	ATAACCGAAT TCTACCATAC CAGGGTCACA GACAATCATC ACTTTTTCAA TCTTGTCAT	1320
	TGTTGTTAGA CTCATGATTG CATTCTCTTC AAAATAAATT TGAGCAGGCA CCTTGAAAAT	1380
5	TTGAGTATTA TTACGTCGTT TAGCAATCGT TTTAATGTTT AATAAATCTG TCGCACTAAC	1440
	ATTATGTGAA ATTGAGTTTC TACCGTAGaA CCACAACCTA ATGTTAAAGA CGGAATCAAT	1500
	TCGTTATACA TATCACCAAT ACCTCCAACC GCTGATGGTG TATTTACAAG TACACGACAA	1560
10	GCTTTCATT C TTAGTCCAAA ATCTTTTTGT AATGTTTCAT CTTCTGTATG GATAACGGCT	1620
	GTGTGTCCTA ATCCACCAAA ATGTAGTGTG TCTTCACAAA TTTGAAATGC TTGTTTTGTA	1680
	GATTGGGCTT T TACTAAGGC TAATACTGGA GATAATTTTT CACGAGATAA CGGATAGTCT	1740
15	GAACCTACAC CGCTAATTTC GGCTATGATA AGTTTTGTAT TTTCGGGGAC AGGTATACCT	1800
	GCTAATTCAG CTATTTCAAC TGCAGATTTA CCGACAATAT CAGGCTTAAT ACCTGTTTTT	1860
20	TGTTCAATTcA TAATTGCATT TTCTAAGCGT TGTAATTCAT CTTTTTTAAC AAAGTATGCT	1920
	TGATGTGCTT TAAATTCATT AGTAACATCT TTATAAATTT CTTTATCAAT GACTACAAC	1980
	TGTTTCAAG CACAAATCAT ACCATTATCA AATGTTTTTG AACCAATGAT ATCATTACT	2040
25	GCACGTTTAA TGTGTGCTGT TTTTCAATG TAAGACGGCA CGTTACCTGG TCCCACACCT	2100
	AATGCCGGTT TGCCAGTTGA ATATGCAGAC TTAACCATGC CCGAACCACC TGTTGCTAGA	2160
	ACTAATGCAA TACCTTTGTG ATTCATTAAT TGTTTTGTG CTTGATAGA AGGCACCTCA	2220
30	ATCCACTGAA TAATATCTTT AGGTGCACCT GCCTTCATTG CCGCTTCTAA TACAACCTCT	2280
	GCTGCACGCT TCGACGATTC TTGTGCACTT GGATGGAATG CAAAAATGAT TGGATTTCT	2340
35	GTCTTAATTG CAATCATCGC TTTAAAAATA GTTGTGACG TAGGATTTGT TGTTGGCGTA	2400
	ACACCACAAA TAACACCAAT TGGTTCCGCT ACATACGTTA ATCCTTTTTT TTTATCTTCA	2460
	CCAATaATCC CTACTGTCTT ATGTCTTTT ATTGAATTCC ATATATATTC AGAAGCGTAT	2520
40	AAATTTTTAA TCGCTTTATC TTCGTATATA CCTCTTCCAG TTTCTTCATG TGCTAATTTT	2580
	GCTAGACCA TATGTTGATC AACAGCTGCT AAGCTCaTTT GATGAACAAT ATGATCAATT	2640
	TCTCTGTG ACTTTTTAGA TAATGCTTCT AATGCTTTTT TCCCTTTGTC AGCTAGAGCA	2700
45	TCAATCATAA TTGCCACTTC TTGTTCTTTC GATCCACGAT TTTCTTTTTT AGGTATAGTT	2760
	AACATATACA ACCACTCCTT TATACTTTGT GAATTAATTC ACAAACATTA TAGTACATGT	2820
	CTCTCAGGAT ATAAAGAAAA TTCTATACAA AAAAGTTTAA TTTCGAATAT TATTTGAACA	2880
50	AATATCAAAT TTTAAAATAA ATGTTTTCAT GAAATCATTG TTATTTCCGT GTTTTATAGAA	2940
	TGATTTTATA ATCATAATTT TTTCAATGAC ATAATTTATT CATAATTATA TATTTAATTC	3000

55

TCCTTGTCGA TACCTATCAA CAGATGTTAC AAATAAAAAC CaCCCGTGTG AACGGGTGGT 3120  
 TTGTTCTGCG gCTATAAGCC TTCCTTACTG GCCaGCCCTA AAAGGGCACT GACAAGTCAG 3180  
 5 CCAACTGCAC TACTATTCCA GCAATCCTAA AGGTTTACTC TTTTTTCTTT CTTTTTTTAT 3240  
 TTTTCTCTCC AGTGAAAGGA TCTAAATATT CTTCATTGA AATTTTCATCT GCAACGATAT 3300  
 CCTCTGTAA TTGATTACGA ATATAATTTT CAATCACTTT TTTATTTCTA CCTACTGTAT 3360  
 10 CCACATAAAA TCCTTTACAC CAAAACCTTC TATTTCCATA TCTATACTTT AAGTTAGCAT 3420  
 GTCTATCAAA TATCATTAAG TTACTTTTCC TTTTAAATAG CCAACAAATG ATGATACCCC 3480  
 AAGTTTGGGT GGTATACTTA CTAACATATG GATATGATCT TTACATACCT CTGCTTCAAT 3540  
 15 TATCTCTACA CCTTTTCTTT CATATAATTG ACGTAATATA ATCCCTATAT CTTTTTTTAT 3600  
 TTTTCCATAT ATCGCTTGTC TTCTGTATTT AGGTGCAAAG ACAATATGTT ACTTACAATT 3660  
 CCATTTGTA TGTGCTAAAC TGTTGTGTC AGATGACATT AAATAGCATC TCCTCGTGTT 3720  
 GATTATTTTG GTTGGCTGAC CAATATTTAC TCTAACATGT AGAGATGCAT TTTTTTGACA 3780  
 ATGGTAGAAC CTTTTCTGGG GAGTGGGACA GAAATGATAT TTTGCAAAA TTTATTTTCGC 3840  
 25 CGTCCCACCC CAACTTGCAT TGTCTGTAGA AATTGGGAAT CCAATTTCTC TTTGTTGGGG 3900  
 CCCC GCCCA ACTCGCATTG CCTGTAGAAT TTCTTTTCGA AATTCTCTGT GTTGGGGCCC 3960  
 CTGACTaGAA TTGAAAAAAG CTaTTaCAA GCGCATT 3997  
 30

## (2) INFORMATION FOR SEQ ID NO: 224:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1391 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

GnGCGAGACA AACACACtTA TTGGTGCCAT TATmCcTAGA ATGaATTCaT ATGCAGTAGA 60  
 TGAAaCAATC AAAGGATTGG CAAAACAATG CAAAAAATAT GAATCaCAAT TAATTTTAAA 120  
 45 TTACACAGGT TTAAATATCG AAGCAGAAAT ACAAGCGCTT GAAACATTAG CACGCaGTAA 180  
 AGTAGATGGT ATTGTTTTAA TGGCTACAGA CATAACAGAG AGACATATTG AAGTCATTAA 240  
 TAAAATGAAT GTACCAATCG TTATTGTTGG TCAACAACAT GAACAACTTC ATAGTATTGT 300  
 50 GCATGATGAT TATAAAGCAG GTCAAATTAT AGGCGAATGG ATTGGTCAAC AGGGATATCA 360  
 ACAAGTTGAA GTGTTTAGTG TAAGTGAAAA AGATATTGCA GTTGGTATAC ATAGAAAACG 420

55

TACTTATGTG GAAGCACAAA AAGATGTTGC AAATGTTTTG GAAAATGTGG AGCAAGTAGA 540  
 TCGCGTTGTT GGAGCAACTG ATACGATTGC ATTAGCTGCC TATAAATATT ATTCTGATAA 600  
 5 AAAAGATGTT ATGAAACCAC ATCAAATATA TGGTTTTGGT GGTGACCCAA TGACACAATT 660  
 AGTGTCTCCA TCGATAAAAA CAATTCATTA TAATTATTTT GAAGCTGGCC AATGCGCGAT 720  
 GGaAGAGATA CAACAGATGC TTAaaaAGCA AGATATGCCA TATAGCGTCA CAGTAGATGT 780  
 10 TAATATTTAG ACGCTGTATT TTTTAAAATA AATGTGGAAC CGATACCATA TAACTATAAA 840  
 TGGATAGGTT AAAAGTTAAA GAACGTAGGT AAAATTGCT ATAATAGAAT ATAAATTGTT 900  
 15 AACAGCATAA ATTATAAAAG GAGGACTGGG TAAATATTAT GACCGAATGG ACTAGAGAAG 960  
 AACGTTATCA ACGAATCGAG GACGTTGATA CTGAGTATTT TAAAACATTA AAACAACAAG 1020  
 TTGATCAATC AAAATTTTCG CAACAATTTT ATATACAACC AGAAACAGGC TTATTAAATG 1080  
 20 ACCCCAACGG ACTTATTTTT TATAAAGGGA AGTATTATGT TTCACATCAA TGGTTCCCAT 1140  
 TAGGCGCAGT ACATGGCTTA AAGTATTGGT ATAACACAC GAGTGATGAC TTAATAAACT 1200  
 TTAAAGCTGA AGGGCCAATT TTAATCCAG ATACTAAATA TGACAGCCAT GGTGTATATA 1260  
 25 GCGGTAGCGC TTTTGAATAT AACGGGCATT TATATTATAT GTACACAGGA AATCATCGAG 1320  
 ATAATCATTG GCAACGACAT GCGAGTACAG ATGATCGCAC GATTGAAAGA AGACGGTTnC 1380  
 AGTTGGnAAA A 1391

## (2) INFORMATION FOR SEQ ID NO: 225:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 930 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

ATTTATTTTA ATGTTTATAT TTTCTAACAC TTTTATATGA TCATAGTAGT AATTGACATT 60  
 TTTCAATTCA AAGACTGGTG TCATCGTATC TCACCTCGCA TTCAACTATA CAACTCCTAG 120  
 45 TAACATATGT AAACAGTAAT GTTTACGACT CAAAATTAGA CAAAATAAAG AGATATGCCC 180  
 CCTTCAAGTT TTATTTATCG CATTTCTTGA AGAGAGCATT ATCATTTTAT TGTTCATAA 240  
 CCTTATTTTT TAATTCTGGG TCAAATTGCT GTTGTTTTAA CATTTCAATT TCAAGTTTAT 300  
 50 ATGGCGGTTT TTTATTTTTT TTATCTTCAC CAACATAAGG TGTTTCTAAG ATTTTCGGAA 360  
 TATCTTTAAA ACTATCATGA TGCACAATGT AATTTAATGC ATCAAAACCA ATGTAACCGA 420

55

GAACAACTTT GATTCTGTCG ACTCCAATGA TTTTATCAAA TTCATTTAAT ACGCCATCAA 540  
 AGTCCTCTTT AACATTATAT CCAGCATCAT GCGTATGACA TGTATCAAAA CATACTGATA 600  
 5 AACGTTCGTT ATTATGAACT CCATCAATAA TACGTGCTAA CTCTTCAAAT GAGCGACCAA 660  
 TCTCTGTACC TTTACCTGCC ATCGTTTCAA GCGCAATACG TACATTATTG TCATTCGTTA 720  
 AAACCTTCATT TAATCCTTCA ATAATCTTAT TAATTCCGGC ATCAACACCA GCTCCAACAT 780  
 10 GCGCACCTGG ATGTAATACa ATATCTTTAG CCCCTATAGC TTGCGTTCTk TCaATTTCTT 840  
 GTTGCAAGAA ATCTACACCA AGATTAAACG TTTCTGGTTT GGTGTAttG CAATaTTaAT 900  
 15 GATGTATGGT GGCATGAACA ACAATATTAG 930

(2) INFORMATION FOR SEQ ID NO: 226:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1984 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

TGACGCACCA ATTTATAACG CAATTGACAA AACAATTAGA TATACCTGTG AAATTTGTAC 60  
 CTGGAAACCA TGATTTATGG GAAGTTGAAA GTATGACTAC GCAAGACATT TGGAATAATT 120  
 30 ATAAGAGTAT GTCACAGTGC TTGGTAGGAA AACCATTAT AGTAAATGAA GAATGGGCAA 180  
 TCATAGGACA TACTGGCTGG TATGATTATA GCTTTGCAGC ACAACGATTT TCATTAGATG 240  
 35 AGTTACAAAA AGGAAAACAT TATGGTGCGA CTTGGCAAGA TAAAGAACGA ATATCTTGGG 300  
 GCATATCAGA TCAAAATTTA TCTAAAATAG CGGCTGAACA AGTGAAGAAA GATATATTAG 360  
 AAGTĀGGAAA TAGACGAGTG ATTTTAGTCA CACATGTTGT GACGCACCCT GATTTCAATTG 420  
 40 TTCCTATGCC GCATCGTATA TTCGATTTT ATAATGCATT TATTGGGACA AGTGATTTCa 480  
 ATCCTTTGTA TGCGATGTC GATATACCAT ATAGTATTAT GGGCCATGTT CATTTTCGTA 540  
 AAAGTGTGAT AGATGATGGC AGATGTTATC TCTGTCCGTG TCTAGGCTAT CCAAGACAAT 600  
 45 GCGGTCAGA AGATATTTAC CAGGAAATAA ATGAGACGAT ACAAATAATA GAAATTTAAA 660  
 ATGCGCAAAC CTGACCCAGT TTGCGCATT TATGTTTTAC ACACGCGAGT AATGTGTTTA 720  
 CTTACGTGTG TTTATTTTGT TGCTGATTTT CAATTGTATA TGAATGTGGT TGCACATAAA 780  
 50 TGCACCTTCT TCCTGGTGAA TTAAAGCTGT ATTCCATTTT CTCTTTACGA ATTTTAATAA 840  
 TTTGTTTGCG ATTTGGAATG ATGGCAGGTA AACTAGGCC ACGACGAATA TGACTCCAAA 900

55

TTGAAACTTG TTTCGCTGGC TTGTTATCAA AGCGGAAAAC ACGTAGTAAT GGTTTAGAAC 1020  
 CAAGATTAGT ATGGTATATT AACACAGGTT GACCTTGATC GATAATACCT TTAAGATCTT 1080  
 5 CTAACGATTT ACCAGTGCCG TCTACGATAT TAGGATTGTA TTTTGTAAA AATGGTACAT 1140  
 ATGCTTCTGG AAATATCGTT TGATGATAAT TGCCAAGCTT AATGAATAAG TGATGTCCAA 1200  
 CATAACCTTT ATGTGGATTG TTCGGATGTG TCGGCCAATG TCTCATAATT TCTGTAGCAG 1260  
 10 GGATATGTTG GTTGTGTAT TGCAACATCA TGGCTGCGGA AACACCTTCA CACCCCATGA 1320  
 CCATAGGGAT AGGAAATAGC TGA CTGATAG GTTTAACTGG TAATATTTTT CGGTTCTATA 1380  
 TATAGTCCTC GCATTGATTC AATAAATATT TAATATAATT ATATAGCGTC AATGCAAAAT 1440  
 15 GTCCTAAACA TATGTTTTAC ATGAGTGAAT AAAATTAATG GAGTGATAAA ATGGAATATC 1500  
 AATTACAACA ACTTGCCTCG TTAACGTTAG TAGGTATTAA AGAAACGTAT GAAAATGGAC 1560  
 20 GACAGGCTCA GCAACATATA GCAGGGTTTT GGCAAAGATG TTATCAAGAG GGAGTAATTG 1620  
 CGGATTTACA GTTAAAAAAT AATGGTGATT TAGCCGGGAT ACTTGGCTTA TGTATACCTG 1680  
 AATTAGACGG TAAGATGTCA TATATGATTG CAGTTACCGG AGATAATAGT GCTGATATTG 1740  
 25 AAAAATATGA TGTCATAACA TTAGCAAGTT CAAAGTATAT GGTATTTGAA GCACAGGGCG 1800  
 CAGTACCTAA AGCAGTTCAA CAAAAATGG AAGAGGTTCA TCACTACATA CATCAATATC 1860  
 AAGCAGATAC GGTAAAATCA GCACCATTTT TTGAGTTGTA TCAGGATGGT GATACTACAA 1920  
 30 GTGGAAAATT AATATTACCA GAAATTTGGG ATnCTGTTA AAGGGGTGAT TGAAATAnGA 1980  
 AnTG 1984

(2) INFORMATION FOR SEQ ID NO: 227:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6373 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

45 GATTCCACGT GTGTAAAAAG AAGTTACAcC TTCAATGATG GTATTTACTA ATTTCTTTAG 60  
 AGATCAAATG GATCGCTTCG GTGAAATTGA TATTATGGTT AATAACATTG CAGAGACAAT 120  
 TAGTAATAAA GGCATCAAAT TATTGCTAAA TGCTGATGAT CCATTTGTGA GTCGTTTGAA 180  
 50 AATCGCAAGT GATACGATTG TGTACTATGG TATGAAAGCA CATGCCCATG AATTTGAACA 240  
 AAGTACGATG AATGAAAGTA GATATTGTCC AACTGTGGT CGCTTATTGC AATACGATTA 300

	AAAATATGAA ATATCAAGTT TTGATGTGGC ACCGTTTTTA TATTTAAATA TCAATGATGA	420
	AAAATATGAT ATGAAAATTG CAGGTGACTT TAACGCTTAT AACGCGTTAC AGCATATACT	480
5	GTTTTAAGAG AGCTAGGGTT AAATGAACAA ACAATTAAAA ATGGCTTTGA AACGTATACA	540
	TCAGACAATG GTCGTATGCA GTACTTTAAA AAAGAACGAA AAGAAGCGAT GATCAATTTA	600
10	GCTAAAAATC CTGCAGGAAT GAATGCAAGT TTATCAGTTG GTGAACAATT AGAAGGCCAA	660
	AAAGTGATG TTATTTGCT AAATGATAAC GCTGCAGATG GTCGAGATAC TTCATGGATT	720
	TATGATGCAG ATTTTGAAAA ATTATCTAAG CAACAAATTG AAGCTATCAT CGTGACAGGT	780
15	ACACGAGCAG AAGAACTTCA ATTGCGATTG AAGTTAGCAG AGGTTGAAGT ACCAATTATA	840
	GTTGAGCGTG ATATTTATAA AGCAACGGCA AAGACTATGG ATTATAAAGG TTTCACAGTT	900
	GCAATACCAA ACTATACATC ATTAGCGCCT ATGCTTGAAC AATTAAACCG TTCGTTTGAA	960
20	GGAGGTCAAT CATAATATGC ATGAATTGAC TATTTATCAT TTTATGTCAG ATAAATTGAA	1020
	TTTATACAGT GATATAGGAA ATATTATTGC TTTAAGACAA CGTGCTAAAA AACGAAATAT	1080
	TAAAGTTAAT GTCGTAGAAA TCAATGAAAC AGAAGGTATT ACCTTTGATG AATGTGATAT	1140
25	TTTCTTTATC GGTGGTGGAA GTGATAGAGA ACAAGCATTG GCAACAAAAG AATTAAGTAA	1200
	AATTAAGACA CCACTTAAAG AAGCGATTGA AGATGGTATG CCGGGATTAA CGATTGTGG	1260
	AGGCTATCAA TTTTAGGGA AAAAATATAT CACGCCTGAT GGTACAGAA TAGAAGGGTT	1320
30	AGGTATTTTA GATTTTATA CTGAATCAAA GACAAACCGA TTAACAGGAG ATATTGTTAT	1380
	CGAAAGTGAT ACTTTTGGAA CTATTGTAGG TTTTGAAAAT CACGGTGGTA GAACATATCA	1440
35	TGATTTCGGT ACACTTGGTC ATGTTACTTT TGGTTATGGT AATAATGATG AAGATAAAAA	1500
	AGAAGGCATT CATTATAAAA ATTTATTAGG TACTTATTTA CATGGACCAA TTTTACCTAA	1560
	AAATTACGAA ATCACTGATT ATCTGTTAGA AAAAGCTTGT GAACGTAAGG GTATTCCGTT	1620
40	TGAGCCTAAA GAAATAGATA ATGAAGCGGA AATACAAGCG AAACAAGTAT TAATAGACAG	1680
	AGCAAATAGA CAGAAGAAAT CTCGTAACT CTGAACATCG CATCAATGGA TTTAATATTG	1740
	ATAAACGATG AAGTTTAGTA ATTAATCATA TATGTATAAA CACACACATT ATTTTGGATG	1800
45	GAAACAACCA AATTGATGTG TGTTTTTTTG TTCTAGTGAA TAATTATTAT ACAATGAGTA	1860
	TCTATCCTAG AATTATCAAT AGTAATGGTG ATTATGCAAC ATGAAAAAAT GAATGATGAA	1920
50	AGGAATTTGA CGATGAAGCC TACTAAAGTG ATATTAAAAG ATGCATCTTA TTTACATAGC	1980
	AAAACATCGA TAACATTTAT TTTAAAAGAT GTAGTTATCG AAGAAGATAA TAAAATTTAT	2040
	TATTTGACA CTAGTGCACT TTCGAAGATC AAGAAGTTAA ATTTGAATTT GCACTCTTTG	2100
55		



	TTATAGAACC	TGATTTACAT	TTTACAATTA	TTGATTTTAA	TCAAGAACTG	CTTTGTATTT	2220
	ATATTGATTT	TGATTCTGGT	TTAAGGCATT	CAAACATGGC	AACAGAATCT	GGTATTTTCAT	2280
5	TAAGGATAAA	TGTTGCTAAA	TCAGATTTTA	CTAAATTTAT	TAATGAATTA	GCCTCTTTAC	2340
	ATTAATGATT	TAAATCTGAT	ATGTAATTAC	AATCAAAAAA	GACAGCCACA	TCCCTCCGTA	2400
	GTTTAGGCGT	GTGGCTATAT	TTGAGTCTGA	ATATTTATGC	TTGTAATTTT	AAAAAGGGAC	2460
10	ATGCTATATA	CGATAAAAAG	AGGCGGGGAC	ATAAATCAAT	GTTCTATGCT	CTACGAAGTT	2520
	ATATTGGCAG	TAGTTGACTG	AACGAAAATG	CGCTTGTAAC	AAGCTTTTTT	CAATTCTAGT	2580
15	CAGGGGCCCC	AACAAAGAGA	AATTGGATTG	CCAATTTCTA	CAGACAATGC	AAGTTGGGGT	2640
	GTGGGCCCCA	ACACAGAGAA	TTTCGAAAAG	AAATTCTACA	GGCAAAGCGA	GTTGGGGTGG	2700
	GACGACGAAA	TAAATTTTAT	GAAAATATCA	TTTCTGTCCC	ACTCCCATGG	TGCCAATTAG	2760
20	CATAAGGTAC	TTAAATTAAG	CATATCTGCT	GTCTAGCAGT	CGATAAATCA	TTAGAACTTC	2820
	GTATAGTATA	TGACTTTTAA	TTTGATTTTC	ACCACTAATT	TCAAGTGCTT	TTATAGTCGA	2880
	ACGTAAAGTT	TCTACAGAAT	CATCTTCTCT	CTTAAAAGAA	CCATCATAAA	ATATATCTTT	2940
25	GATGCTACTA	CTAATTTTAA	GCAATGCCAT	TTTTTCGTCA	CCTGAAAAGT	TAACACGAGT	3000
	ATTTTTAGGC	AAGTAAATGA	TATTTGATAA	ATGAGTGATA	AACAAACGAT	TCGTATATGC	3060
	ACGTTTAGTT	AATTGATTGA	GTAATTTCCA	ATCACATTCT	TTTTTCTTAT	GATAGCTTAA	3120
30	TTCATCACGT	TGATAACTTA	TTAACGTTTC	AACCTGATTA	TTTAAATTGA	AAATATTTTT	3180
	ATATGCTTTT	TCGCTTTTAT	CAGATTGCAG	TCTTGATAAG	ATAAGTTCTT	GGCAGCGATT	3240
35	GTAAAATAAT	TTATACATCA	AGGCATCTGT	CTTACTTAAT	TTTTCTTCGA	CCTGACCATA	3300
	ATACTTAGGT	GGAAACACCA	TGAAGTTAAT	TAAACCTGAT	GTCACGAGTC	CAATAATTGC	3360
	TGTCATATGTT	CGAGACAAAA	AGTTGAATAT	GTAGGCATCA	TGAATACCTG	GAATCATAGC	3420
40	TAATGATGTT	AGTACAGCGA	CATTCGTACC	AACCTGCAAT	TTGAGTTTTG	TACAGAATAA	3480
	AATCGTGAAC	GTTGCACTCA	ATGCATATGT	AAAAGGTGAT	TGATCGCCGA	ATAAATATGT	3540
	AAATAATACT	GCAAAGCCTG	CACCAATTAC	CGTAGCAGGT	AATCTACGAT	AACCTTTAAT	3600
45	AAGTGATGCC	TTGGCAGTTG	GTTCAATTGT	GACTACAGCT	GTTAAAATGG	CATAGATGGG	3660
	TGTTAAATCT	AGTGCCATAC	AAAAGACAGC	TGTTAAAAAA	ATGGCAATAC	CAGTTTTAAT	3720
	TGTTCTGGCA	CCAATTAAAT	GTTTATACCA	TTGATCGTTC	ATTTTTTAAC	CTCTAATCAT	3780
50	CGTAAAATCT	TAGCGAGCGC	TTTATAATAA	TAGTATCGTA	CATTGGAAAA	GTTTCATGTAT	3840
	GTAAAATATT	TGAAATAATC	ATACATAAGC	ATTACTTTGA	TTTTCATATA	CATTAATCAA	3900
55							

	CAAGCATT	TTT TCAATTATAG	TCCGGGGCCC	CAACATAGAG	AATTTCAAAA	AAGAAATTCT	4020
	ACAGGCAATG	CAGGTTGGCG	GGGCCCAAC	ACAGAAGCTG	ACGAAAAGTC	AGCTTACgAT	4080
5	AATGTGCAGG	TTGGCGGGGC	CCCAACATAG	AGAAATTGGA	TCTACAATTT	CTACAGGCAA	4140
	TGCAAGTTGG	GGTACAACGA	TAAAGAAATA	TTTTTCTTT	ATCACACTAT	GTCTCACTCA	4200
	CTTTCCAAAA	TACTAAAGTA	ACATCTTTAG	TATATCAAAG	AATTTTGTCT	ATAATAAGTT	4260
10	ATAATTATAT	AAAAAAGGAA	CGGGATAAAA	TGATTGTAAA	AACAGAAGAA	GAATTACAAG	4320
	CGTTAAAGA	AATTGGATAC	ATATGCGCTA	AAGTGCGCAA	TACAAATGCAA	GCTGCAACCA	4380
15	AACCAGGTAT	CACTACGAAA	GAGCTTGATA	ATATTGCGAA	AGAGTTATTT	GAAGAATACG	4440
	GTGCTATTTT	TGCGCCAATT	CATGATGAAA	ATTTTCCTGG	TCAAACGTGT	ATTAGTGTCA	4500
	ATGAAGAGGT	GGCACATGGG	ATTCCAAGTA	AGCGTGTCT	TCGTGAAGGA	GATTTAGTAA	4560
20	ATATTGATGT	ATCGGCTTTG	AAGAATGGCT	ATTATGCAGA	TACAGGCATT	TCATTGTGTCG	4620
	TTGGAGAATC	AGATGATCCA	ATGAAACAAA	AAGTATGTGA	CGTAGCAACG	ATGGCATTG	4680
	AGAATGCAAT	TGCAAAAGTA	AAACCGGTA	CTAAGTTAAG	TAACATTGGT	AAAGCGGTGC	4740
25	ATAATACAGC	TAGACAAAAT	GATTTGAAAG	TCATTAAAAA	CTTAACAGGT	CATGGTGTG	4800
	GTTTATCATT	ACATGAAGCA	CCAGCACATG	TACTTAATTA	CTTTGATCCA	AAAGACAAAA	4860
30	CATTATTAAC	TGAAGGTATG	GTATTAGCTA	TTGAACCGTT	TATCTCATCA	AATGCATCAT	4920
	TTGTTACAGA	AGGTAAAAAT	GAATGGGCTT	TTGAAACGAG	CGATAAAAGT	TTTGTTGCTC	4980
	AAATTGAGCA	TACGGTTATC	GTGACTAAGG	ATGGTCCGAT	TTTAACGACA	AAGATTGAAG	5040
35	AAGAATAGTT	CAACATATAC	TAAGACTAAA	GTATGAACAT	CATTTAGTTC	CGGAGCCTAT	5100
	TCATATTGGT	TTCGGAACTG	TTTTATAATA	ATTAAGAACA	CAATCAATGC	GTCATTTCAA	5160
	AAATATGTTG	TAACAAAGTA	GTTTTTAAGC	AAACATATCA	TCGACATCAA	CGAAGATACA	5220
40	TAGCGCATTT	GGTATTTTAA	AAC TTATTAT	AAAAGGTGAT	AGTTATGAAC	TATGTTGAAC	5280
	GTTATATTGA	ACAGTTTGTG	AGAGCAACAG	TAAGAAATAA	TATCAAGCAC	TACCTTTTAA	5340
	TGCTAGATGA	AAAAATGAAA	AATTTAGATG	ATTATATGCG	TTATTTAATT	ACTAAAAAAG	5400
45	AACAACCTTAG	CAAGTTAATT	GACAGTCTAA	TGCTAACATT	AGAAAATAAA	TATATTGATA	5460
	TTGCTGAAGC	ATTTCAAATT	CAATGTGCAA	GAGAAATCAA	TAATCAAGAA	ATTGAAAATA	5520
50	TTAAATCAGA	GTTGAATAAA	GTTGAAGCAT	ATTATGCACA	AATTGAAACT	CAAATTCAAC	5580
	AAACTTCAAC	TGAAAAAATA	GCAACAGAAA	AAACATCGTA	TCTAATAAAT	TATATGAACG	5640
55	CTGTGGCATA	GAAAGGCGGC	GAAACATGAC	ACACAAATAT	ATATCAACGC	AAATGTTGAT	5700

CTTTTTACTC GTTCTATTAT TGGGATGTGT ATTAGTTTAT GTAGGATATC TTTATTTTCA 5820  
 TAAATACGT GGCCTTTTGG CGTTTTGGAT AGGCGCGCTA TTAATTGCAT TCACATTATT 5880  
 5 GTCTAATAAG TATACAATCA TCATCTTGTT CGTCTTTTTA TTATTACTTA TTGTGCGTTA 5940  
 TTTAATACAC AAGTTTAAAC CAAAAAAGT AGTTGCGACG GATGAGGTTA TGACTTCACC 6000  
 ATCTTTTATT AAACAAAAGT GGTTTGGTGA GCAACGTACA CCAGTTTATG TATATAAGTG 6060  
 10 GGAAGATGTA CAAATTCAAC ATGGAATTGG CGACCTACAT ATTGACTTAA CAAAAGCTGC 6120  
 AAATATTAAG GAAAATAATA CCATTGTTGT TAGACACATT TTAGGTAAAG TGCAGGTTAT 6180  
 15 ATTGCCGGTT AATTACAATA TTAATTTACA TGTAGCTGCT TTTTATGGAA GTACTTACGT 6240  
 GAATGAAAAA TCATATAAAG TTGAAAATAA CAATATTCAT ATTGAAGAAA TGATGAAACC 6300  
 GGATAACTAT ACAGTTAATA TCTACGTATC AACGTTTATC GGAGACGTAG AGGTGATTyA 6360  
 20 TCGATGAAYC ACT 6373

## (2) INFORMATION FOR SEQ ID NO: 228:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 4488 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

ATAGnGAAAG CGTTTTACAC TTAATAACTC CCTCTTAAAT GCATCCAGGT TCTATGTAGT 60  
 35 AAATCATGAA nATAACATAT AAATnTAGAG GAGATTTACC TTTGAATACA GAGAACAACA 120  
 AGAATCAAAA CCAATCTGTT AAAAATTCTG AAAGaCGCGG CATGTTAAAA GGATGCGGCG 180  
 GTTGCCTTAT TTCTTTTATT TTATTAATAA TCTTATTATC AGCCTGTTCA ATGATGTTTA 240  
 40 GTAATAATGA CAATTCCACT AATAATCAAT CATCAAAAAC GCAATTAACT CAAAAGATG 300  
 AAAATAAAAA TGAAGATAAG CCTGAGGAAA AATCAGAAAC AGCAACAGAT GAGGATTTAC 360  
 AATCAACCGA AGAAGTACCT GCAAATGAAA ATACTGAAAA TAATCAACAT GAAATTGATG 420  
 45 AAATAACAAC AAAAGATCAA TCAGACGATG ATATTAACAC ACCAAACGTT GCAGAAGATA 480  
 AATCACAAGA CGACTTGAAA GATGATTTAA AAGAAAAGCA ACAATCAAGT AACCATCATC 540  
 50 AATCCACGCA ACCTAAGACC TCACCATCAA CTGAAACAAA CACGCAACAA TCATTTGCTA 600  
 ATTGTAAGCA ACTTAGACAA GTATATCCGA ATGGTGTGAC TGCCGATCAT CCAGCATATC 660  
 GACCACATTT AGATAGAGAT AAAGATAAAC GTGCATGTGA ACCTGATAAA TATTAAACAA 720

55

	GGGAGATTTT TTAGGCATGA GCAATCAATT CAAAAGCGAA GAAGAGCGAA GACAATGGGA	840
	ACAATTCCAA GCTTTCCAAA ATCAACAAAA CCAACAGAAC CAGCAATACG GACAAAAGAA	900
5	ATCTAAAAAA GGATGGTTCT GGGGCTGTGG TGGTTGTCTA GTATTATTTA TTTTAATTAT	960
	CATCGGTATT TCAGCTTGTA CAGCTGGTAT TACAGGTAAC CTTGGCGGAA ATAGTTCTAA	1020
	AGAAACGAAC AAAACCCATA AAATCGGTGa AACTGTTAAA AATGGCGACC TTGAAGTCAC	1080
10	TGTAAATTCA GTGGAACTA TGAAATCTGT AGGACCATCT CTTGCACCAA CAAACGCTAA	1140
	AGGTATATTT GTCGTTGCTG ATGTGACGAT TAAAAACAAA GGTAAAGAAG CGTTAACAA	1200
15	TGATAGTTCA ATGTTTAAGC TAAAATCCGG TGATAAAACA TTTGAAGCAG ATAATACAGG	1260
	TTCAATGTCT GCTAATCAAA GTGACAATGG TAGTATAGAA AATTCATTTT TCTTACAGCG	1320
	TATAAATCCA GATAGCACTG CTCAAGGTAA AATTGTTTcG ATGTGTCAGA AAACATAGCC	1380
20	AACGCAAAaG ATAAAAAATT AGAAGTTATT TCTAGTTTAT TTAGCGTCAA GAAGATTACA	1440
	TTTGATTTAT CCGATGCTAA AAAAACATCA AAAGCTAAAA AAGACAAGCA AGATACAGAA	1500
	GTAGCTGTTG CGAGTTCAAA TAGCGATAAT GTAAGTTATG AAGCTTCGGC TACTACACCT	1560
25	GcTACAACTT CTAGTGCGGA TACTGATTCT GAAGATAGCG AAAAGTCTAG TAAAGATGAG	1620
	GATAAGCAGA ATGCGTCTAA AAgTGATAAA TCTAGTGTAG AAAAAAGTGA ATCTAATGAG	1680
	GAAACTGCTC CTGTAGAGCC CATGCCCCAT AGCAAACCTA CCACTAGTGA aGCACCACCT	1740
30	AGCCAAAATA TTCACAaTGa AGATAGCmTG TACGACGCTT CAACAGAATA AAATtnyCAG	1800
	tAGCTCGGCT ACCCTTCTTT TACGGAAAAA TTAATTATAC ATAATCaAAT CaAGGAGATA	1860
35	AAAAAATGAA ATTCAAAGCT ATCGTTGCAA TCACATTATC aTTGTCACTA TTAAGTGCCT	1920
	GTGGTGCTAA TCAACATAAA GAAAATAGTA GTAAATCAAA TGACACTAAT AAAAAGACGC	1980
	AACA <sup>̄</sup> ACTGA CAACACTACA CAGTCAAATA CAGAAAAGCA AATGACACCA CAAGAAGCCG	2040
40	AAGATATAGT TCGAAACGAT TACAAAGCAA GAGGCGTTAA TGAATATCAA ACATTAAATT	2100
	ATAAAACAAA TCTTGAACGA AGCAATGAAC ATGAATATTA TGTTGAACAT CTAGTCCGCG	2160
	ATGCAGTTGG CACACCAITTA AAACGTTGTG CTATTGTTAA TCGACACAAT GGCACAATTA	2220
45	TTAATATTTT TGATGATATG TCAGAAAAAG ACAAAGAAGA ATTTGAAGCA TTTAAAAAGA	2280
	GAAGCCCTAA ATACAATCCA GGTATGAACA ATCATGATGA AACAGATGGT GAGTCAGAAG	2340
	ACATTCAACA TCATGACATT GATAATAACA AAGCAATTCA AAATGACATA CCAGATCAAA	2400
50	AAGTCGACGA TAAAAATGAT AAAAATGCTG TTAATAAAGA AGAAAAACAT GATAATGGGG	2460
	CAAATAATTC TGAAGAACT AAAGTTAAAT AATGGCATAC TTTGATTAAT CGTAATTTTT	2520
55		

	ACTATGCATG GTCTTTTAA TCAACTTAAA CTCGGCATT TTTCAATCGA AAACGCAGAG	2640
	CATACGCTTT TTACACCTTA TATGTTGGAA ACGCTCTCTT CCCTAGGCGT GAAAGACAGC	2700
5	ATTGTCGATT TAATTCATAA AGGGACTGAA TTAGAAGACT TTGCGGCATT TAATTTATCA	2760
	ATTGAAGACA CAGTTACAGT CTGTTTACAA AGAACTGAAG AACTATTAAA ACAATACAAA	2820
	AATGTGGAAT TCAATGACAA AATATTAATC AATTGGCGTA TTATACAAGA GAAATAGACA	2880
10	TATAAAAGTC GAATGTAACt ACGTGAGTAT TGATTTTATT CTTTGTAAT TACAAGCATT	2940
	TCATATTATA AAGTTTGAAA AGAGGTATAT TGAAATGGAG AAAAAATGAAT ATATAGCTAA	3000
15	ATATAATGAA TATAGTCAAT TATTAGACGC TACATACTCG CAAGCTGTAG CATmCCTTTT	3060
	AAGtAAATaT GGCCTGTAA CCGATGATTA TTATAAAGaa AAATCATACA CGCGATTTTT	3120
	AAAtGGAGnA ATCAAAAGTA TTTCAAAAGG AAAATACACT AGAGCTAGCG AaGGATTATA	3180
20	TTGCCATCAT ATAAGCGAGG ACAAATTCCA AAATCTATCT GATCTAAGAT TCATtTCCAA	3240
	ATTTAAGTAC TCATACGACG TTCAAAAGAA AGAAACTTA GTGTACTGTG ATCTAATCGA	3300
	GCATTTAATT TTACATGCAA TTATTACAAA AGAATCCCAT GGCCAATTTG GTGTAGCTGG	3360
25	ATTATGTCAA ATGATCAAAC CAACAGTCAT TGATTGGTAC ATTGGCGAAT ATAATCCAAA	3420
	ACCAGCATGG ATGCAAGCCA CCAAAGCAGC TGCCTATTTG CCTGGAATAT TAGTAGAGAA	3480
	ATTACTCATT AAAATTGACG ATATGTTAAA AGGAATAGAA ATATAAGATT TCCTTGAGTC	3540
30	TAGATAAATG ATTAATGTAG ATTTATTTTT TGCTGTTGAG ATTTTGTTAT AGATGTTTAA	3600
	ACCTGTAATT AAATATATTT TATAAAATAG ACCACGCATA CCTATCTATA AACGGrCAAT	3660
35	GTTTATAAAT GAGTTTGCAT GGtCTTGAAT TGTATTAAAT TTCTTTTGGT TTTAATAAAT	3720
	CGACTAGATT TTCACAATAT TTATCAAATA TGTATTCCTA AATTATACAG CCTTAATCCA	3780
	GCAgCTACTT TCGAAACTTC CAACTTAGTT GATATAAGGT TCAATAGTTT GTTTCGTTCT	3840
40	TTTTCAGATA AACCAGAACT TAAATTGATA TTATTGACTT CATAAAAATT ATAGACTAAT	3900
	GCCTCTATTT GCTTTTTAGG CATAAGTAAG TCGACTGAAA ACTGATTTAC GTCGCTTTCA	3960
	TAAATCATT CATGTAAATT CTTTAGACTA TTATCGTTAC TATCTCTCAT TAAGTCTGTA	4020
45	TTTTTAAATA AATAACGGCC CAATTCACGA GCTATTGAAA ATCTTGTTAT ATTAATCGAG	4080
	TGATTATTAT TAATATAGAT TGTTCTTCCA CTTAAATAAC CCGAAGTATT ACCCTCCATT	4140
50	TTAATATATC TAACATTTAA ATTAAGTTGA AATAATAGCT TGTCTATGTC AATAGCAAAG	4200
	TGTTTCAGAAG TAATAAAAAG TTGATCCATT TTGTCTTTA TAAATGCCTG AAATAATCGA	4260
	ACTATTTTTG ATTCTAAAAT ATCTTCATAA TGAACCTTCT CAATAACTTT CAATTGATTC	4320
55		

AACATTATTA AAATAAAACC CCTCTACTAC TATATGTAAC GAAGGGACAT GATTTCAAAA 4440  
 TAAAATACCT TTTTATATAA TnTATTATAA TATCCCCCAG TATACnAC 4488

5 (2) INFORMATION FOR SEQ ID NO: 229:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 846 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

TATGGCGCCA TATTAGTTGT AACTGGTTTA AGAGGTCCaA GAAaTATCaA ATAAAGTTGT 60  
 tCCTGGGCTT GGTACTGTTA TCTCaATATT GmwTGCaTTT GGTGGTCTAG CTTTTAATAT 120  
 20 TGGTAATATT GCTGGTGCCG GTTTAGGTTT AAATGCAATT TTTGGATTAG ATGTAAATG 180  
 GGGCGCAGCT ATTACTGCAA TCTTTGCAAT ATTAATCTTT GTAAGTAAAA GTGGCCAAAA 240  
 AATTATGGAC GTTGTTTCAA TGATTCTTGG TATTGTGATG ATTTTAGTTG TGGCATATGT 300  
 25 GATGTTTGTT TCTAATCCAC CTTATGGTGA TGCTTTTGTG CATAATTG CGCCAGAACA 360  
 TCCAATGAAA TTAGTCTTGC CCATCATTAC GTTAGTTGGT GGAActGTar GTGGTTATAT 420  
 TACCTTTGCA GGTGCACATC GTATATTAGA CTCTGGCATT AAAGGTAAGC AATATTTACC 480  
 30 ATTTGTAAAT CAATCAGCAA TTGCTGGTAT TTAACTACA GGTATTATGA GAACGTTACT 540  
 ATTCCTAGCG GTATTAGGAG TTGTTGTAAC AGGTGTGACA CTAAGTTCTG AAAATCCACC 600  
 35 AGCGTCAGTT TTGAACACG CAATTGGACC AATTGGAAAG AATATTTTGT GTATTGTGTT 660  
 ATTTGCTGCA GCTATGTCAT CAGTAATTGG CTCAGCATAA ACAAGCGCAA CATTTTTTAAA 720  
 AACACTTCAT AAATCACTTA ACGAAAGAAG TAATTTAATT GTGATTGTGT TTATCGTTAT 780  
 40 TTCAACAATG ATTTTCTTAT TTATTGGAAA ACCAATCAGC CTTTTAATTA TAGCAGGCGC 840  
 GATAAA 846

(2) INFORMATION FOR SEQ ID NO: 230:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2072 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

55

	TCTTTTAAAGGTTACTAAT ATTTCTTTAG TGAAAATTGA ATCACGGTCG TTTATTGGTG	120
	CCTTGAGTAT ATTATTATAG ACGGAATCTG ATCTAATAAT ATTGATTTTA TACATGATAA	180
5	ACCTCCTTAT GTTGTcAGCA TAAAGGATAA CGTAACGTGA TTTTCAAGCA GTAATTGTAA	240
	CTAATTGAmA AAAATTAAGA AAAGTATGTG AGTGTTCCCTA AwTAATATGa TTAAAATGAT	300
	GGCGAATAAG TGTCTaAAAG CATCTTAAAG GGACATTGTA TAGGGTAAAT CACTTCATAA	360
10	ATAAGGGaAA ATCCTTATGT TCACTTTTTTC ACAATCATnA TAAAATATAT ATGTAGTCAA	420
	TACTTTGTCT ATATTGAATG TTTTCATATA AATGAAAGCA TTTTAAATA ACATTGACCT	480
15	CTAATATATA GGCAGAGTAT TGATATCTAT TAAAAAATAA ATGATTTTGA TGAAGGTGAA	540
	ACGTATGTAC AAAACAAAAG GTGGCTTTCA ACTTACATTA CAAACATTAA GTTTAGTGGT	600
	TGGGTTTATG GCTTGGAGTA TAATTGCGCC ATTAATGCCC TTTATTAAAC AAGATGTCAA	660
20	TGTTACTGAA GGTCAAATAT CAATCATTTT AGCGATACCA GTTATTTTGG GATCGGTGCT	720
	CCGTGTGCCA TTTGGTTATT TAACAAACAT TGTTGGCGCT AAATGGGTAT TCTTTACTAG	780
	TTTTATCGTA TTGTTATTCC CGATATTTTT CTTAAGCCAA GCACAAACAC CGGGTATGTT	840
25	AATGGCTTCA GGATTTTCC TTGGTGTAGG TGGTGCAATT TTCTCAGTTG GTGTTACATC	900
	AGTTCCTAAA TATTTCCCTA AAGAAAAAGT AGGTCTAGCA AATGGTATTT ATGGTATGGG	960
30	AAATATCGGT ACAGCAGTTT CTTCATTTTT AGCACCACCG ATAGCGGGTA TTATTGGTTG	1020
	GCAAACAACA GTTAGAAGTT ACTTAATTAT TATCGCTTTA TTTGCATTAA TTATGTTTAT	1080
	TTTTGGTGAC ACACAAGAAC GTAAAATTAA AGTACCATTA ATGGCtCAAA TGAAAmCATT	1140
35	ATCTAAAAAC TACAAATTAT ATTACTTAAG TTATTGGTAT TTTATTACTT TTGGTGCTTT	1200
	TGTAGCATTT GGTATTTTCT TACCTAACTA CTTAGTAAAT CATTTTGGAA TTGACAAAGT	1260
	AGAT <del>S</del> CTGGT ATTCGATCAG GTGTATTCAT TGCGCTGGCA ACATTCTTAA GACCAATAGG	1320
40	TGGCATTTTA GGTGATAAAT TTAATGCAGT TAAAGTATTG ATGATTGATT TTGTTGTTAT	1380
	GATTATCGGT GCCATTATTT TAGGTATTTT AGACCATATC GCATTATTCA CTGTAGGCTG	1440
	TTTAACAATA AGTATTTGTG CAGGTATTGG TAACGGCTTA ATCTTCAAAT TAGTACCATC	1500
45	ATACTTCTTA AATGAAGCGG GATCCGCAAA TGGTATCGTA TCAATGATGG GTGGTTTAGG	1560
	AGGATTCTTC CCACCACTAG TAATCACGTA CGTAGCTAAT TTAACAGGAT CAAGTCATTT	1620
50	AGCATTTATT TTCTTAGCGG TATTnGGAnG TATTGCATTA TTTACCATGC GTCATTTATA	1680
	CCAAAAAGAA TATGGCTCAT TGAAaAACGG TTGATATGTA ATACATGCCA TTcATTTAGT	1740
55	TAAATACAAA GCCTTaATAT CATGCGCAAT ATTCGTAGCA TGACATTAAG GCTTTAGTAG	1800

CTTGGTTTGA TTTTAGGCAA GGTAATGGTT AATAACCCAT TTTCAAAACT AGCAGTAATA 1920  
 TGTGTCTTAT CAACAGCTTC AAAATCAAAT TGACGCATTA ATGATTGCGAA GTTACGCTCA 1980  
 5 TCTAAATGA GTTGTTGAGA TTTGTATTTT GCGCTTCTAG TAGCTTGAAT AGTGAGCGWA 2040  
 TTAttATTGA AATCgATACT AATAtCTccc TG 2072

(2) INFORMATION FOR SEQ ID NO: 231:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3159 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

CGTCTTCTCT TGGTTATTCT CTGTGTTTTG TCTTTGTTCA AATTCGATTT TTTGTTGTTT 60  
 GAATGAATGT AATACTTTAT TTTTTTTTGG CACATAATCC AAGTnATTTT TAGGAATTAA 120  
 TATACGGTCT TTAAATGCTT CTTGTATTTT GCTCACAATC AATTGGTATA GTTGCTCTTC 180  
 25 TTTTGATAAA CGCACTTCTA GTTTTGTGG ATGAACATTT ACGTCTACTA AGATTGGATC 240  
 CATTTCAATA TTAATATAAC AAATCGGGAA CCTACCTATT GTTAAGAGTG TATGATAGCC 300  
 TTCTAAAATC GCTTTATTTA GCATAAAGTT TTTAATGTAT CGTCCATTAA TAAAAATAGA 360  
 30 AATATAGTGC TTATTACTTC TAGAATGTTT AGGCTTTGCA ACAAACCTT CAATGTGATA 420  
 ATCACTTGTA TCTCCAGATA TATGTACTAA ATCTCGTGCA ACTTTCATCC CATAAATCTC 480  
 35 TGCCATCACT TCATTAGTTC GTCCTGAACC ATTTGTACTT AACATTGTTT TGCCATCTGA 540  
 AATGAGTGCT ATTCGAATGT CCGGATGGCT CATTGCCATT CTGTTGACAA TATCTGTTAT 600  
 TTTACCTAGT TCAGTGATA AACTTTTAAT ATATTTTAAA CGTGCTGGTG TATTATAAAA 660  
 40 TAATGATTCT ACAAGTATAT CTGTTCCCTT TTTGCTTTT GCAGGCTTAT GATTTAATAT 720  
 TTCACCATTT TCTACATATA TTTCATTTCC ATTAGCATT TCCGTGCAAG TCTTCAATGT 780  
 TACTTTAGCA ACTGATGAAA TACTGGCTAG TGCTTCACCA CGGAATCCTA ATGTCCTAAT 840  
 45 ATGAAATAAA TCTTCATCTT GATCTAATTT ACTAGTCGCA TGTCTATGAA ATACTAATCC 900  
 TAAGTCTTCC GCTTCAATTC CGCTTCCATT ATCGACTACG CGAATAGATT GGACGCCAGA 960  
 50 tTCCTCTACT TCAATGCTTA TTTCTGTAGC GCCTGCATCT ATAGCATTTT CCAACAGTTC 1020  
 TTTCAACA GAACCTGGTC TTCAACTAC TTCACCTGCT GCGATTTTAT TTGCTAATGA 1080  
 GGTTTGGAGT TCTTTAATTT TCCCCATTTT GCAACACCTC TATTTTAATT GATTTTGTA 1140



	TTGTAGTTCA ATCTCGCTTT TTTGATCATT TTCAAACAAA TcAAATGATG CyTGTTCAAA	1260
	GTCTTTTTGA GATAAAGTAT CaGTTGTTTC TTCaACACTT aAGTTTAAAT TTTCTTGATT	1320
5	AATTTCCAGG TTCATTTTCG ACCATTTTTTA AATTTGATAT CGATGATtTT TCACCAGCAG	1380
	ACGCTTCAAA CTCGCTTAGA ATCACTTGTG CTCTGCTAAT AACTTTTTCA GGTAAATCAG	1440
10	CTAATTTTCGC AACTTGAATA CCATAAtATC GTCAACTGCA CCATCTTTGA CTTTATGCAA	1500
	GAATATAAGT TCACCTTTAT ATTCAATTAGC AGCGACGTGA ACATTTTTTA GACTTGGTAA	1560
	TGCTTGATCT AATGTTGTCA ATTCATGATA ATGTGTTGAA AATAACGTTT TAGCATGTGA	1620
15	TGTTTCAGCT ACATACTCTA TCATTGCCTG CGCTAAAGCT AAACCGTCAT ACGTTGAAGT	1680
	ACCACGTCCA ATTCATCGA AAATAATCAA ACTATCCTCT GTTGCAATAG TTAATGCCTT	1740
	TTGTGCTTCT AGCATTTCTA CCATAAACGT ACTCTTACCT GAAACCAAAT CATCTGCCGC	1800
20	ACCTATTCTA GTGAATATTT GATCAAATAT AGGTAACACT GCCTCTTTAC AAGGGACATA	1860
	AGCTCCCAT TGGGCCATTA TACTAATTAT GGCAACTTGT CTCATATATG TCGATTTACC	1920
	AGACATATTC GGACCTGTAA TTAAATATAT AAATGTTTCA TTATCTAATC GACAATTATT	1980
25	AGGCACATAG TCATTATAAT CCATTACTCT TTCCACTACT GGGTGCCTAG ATTCCACTAA	2040
	TTCTAATGTT TTATTTTCAC TAAATGAAGG CCTAGTGTA TATATTTTTT GAGCAATTTT	2100
30	TGCAAAGCTC TGTAACAAT CTAGCTCTGA AATAATTTTA GCTTGTTGTT GTAAACGTTT	2160
	AGTATATTTT TTAACCTCTT CACGTAGCTG AACAAATAAT TGATATTCTA ATTCGATGGC	2220
	TTTGCTTCC GCACCTAAAA TGATATCTTC TTTTCTTTA AGTTCATCAG TTATAAAACG	2280
35	TTCAGCATTC GATAACGTTT GCTTCCTCAT ATAACCAAAT TCACTTGGTT CAAAATTTTG	2340
	CAAGTTGGCA CGTGTTATTT CTATAAAATA ACCAAACACT TTATTAAAGC TTATTTTCAA	2400
	TGATTTTATT CCTGTACGTT GTCTTTCTTT GGCTTGTAAT TCTGCTAACC ATGTTTTTCC	2460
40	GTTTTTTGAA GCTTCAAGAT ATTCATCTAA TTGCGTATTA AAACCAACTT TGAATAGTCC	2520
	GCCATCTTTA ACTGAAATTG GTGGTTCTTC TACTAACTC TGTTCTAATA TATCAAGTAA	2580
	ATCATCaAGG GGTCTAGTT GATTAACTTG TACAAGAGTA TTCTGATTCA TAGAATTTAG	2640
45	TAATGCTTTA ATATTGGTA TTTCAGAAAT GGAATGTTTA AGTTGAATTA AATCTCTCGC	2700
	ATTAACATTT CCGTAACTAA CACGCCCAAC AAGACGTTCA ATATCATACA CTTGATTAAG	2760
50	ATATGTTCTT AAGGTGTCTC TTTCTATGAA ATGAGCACTA AATTCATCAA CGATATCTAA	2820
	TCGTGCTTCA ATTTGTTCTT TACTTATTAG TGGTCTATCT ATCCATTGTT TTAAGCGGCG	2880
55	TGCTCCCAT GGTGTTTTCG TTTCGTCCAT TAGCCAAAGT AGCGTTCCTT TTTTGTATT	2940

ATCTATAGCT GCATATTGAA CAACATCCTC GATATGCGAT AAATCACGTT TTTGTGTATG 3060  
 ATGAATATAA TCTAGCAATA ATTGTGTCCG TTGATACaTT AATTTATGTT CAGTTTGATT 3120  
 5 CACACTATAG ATTtCTGATG ATAACGTTTC CCTGACTGT 3159

## (2) INFORMATION FOR SEQ ID NO: 232:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1238 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

TAAGCGAGAA GCAATTGGTG TTATGTATGC TAGTGATAAA CCAACAGGTG AAAGTACAAG 60  
 20 GTCATTTGCT GTTTATTTCT CTCCTGAAAT TAAGAAATTT ATTGCAGATA ATTTAGATAA 120  
 ATAAATCATC CATCCATACA TTGATAAATG ATTTTtYAGAA ATTAACAACA AAATCAACAA 180  
 TTTTAAACAT CTCTGTGATT CTATTTATTC GAAATGATT AAAAAATAAA ACTTCAAAAA 240  
 25 CCTAACCTTA TATTTATACG AATACTTAGA GGAGCACAAA AATGAATAAA AATATAATCA 300  
 TCAAAAGTAT TGCAGCATTG ACGATTTTAA CATCAGTGAC TGGCGTCGGC ACAACAGTGG 360  
 TTGAGGGTAT TCAACAAACG GCTAAAGCTG AACATAATGT GAAACTAATC AAAAATACTA 420  
 30 ATGTAGCACC ATACAATGGT GTCGTTTTCA TAGGATCTGG AACAGGTTTC ATTGTCCGTA 480  
 AAAATACAAT TGTTACCAAC AAGCATGTCG TTGCAGGTAT GGAAATTGGT GCACATATTA 540  
 TAGCGCATCC CAATGGTGAA TATAATAATG GCGGATTTTA TAAAGTTAAA AAAATTGTCC 600  
 35 GTTATTCAGG TCAAGAAGAT ATTGCCATTC TACATGTGGA AGATAAAGCT GTTCATCCAA 660  
 AAAACAGGAA TTTTAAAGAT TACACAGGCA TTTTAAAAAT AGCATCAGAA GCTAAAGAAA 720  
 40 ATGAACGCAT TTCAATTGTT GGCTATCCAG AACCATATAT AAATAAAATTT CAAATGTATG 780  
 AGTCAACAGG AAAAGTGCTG TCAGTTAAAG GCAACATGAT TATTACTGAT GCTTTCGTAG 840  
 AACCAGGCAA CTCAGGTTCA GCTGTATTTA ACAGTAAATA CGAaGTtGTA GGTGTTCACT 900  
 45 TTGGTGGAAG CGGCCCTGGA AATAAAAGTA CAAAAGGATA TGGTGTtTAT TTCTCTCCTG 960  
 AAATTAAGAA ATTCATTGCA GATAACACAG ATAAATAAAT CCTTACATAG ATAAATGATT 1020  
 TTAAAAATTA ACAACAAACT CAACaATTCA AATCATCTCT GTGATTCCAT TTATTCGAAA 1080  
 50 TGATTAAAAA AAATAAAACT TCAAAAAGCT AACATTATAA TTATACAAAT ACTTAGAGGA 1140  
 GCAGAAAAAT GaATAAAAAAT ATAATCATCA AAAGTATTGC AGCATTGGAC GATTTTTAAC 1200

## (2) INFORMATION FOR SEQ ID NO: 233:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6444 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

	TGATAAGTCA TTTAAATTGT CACCTATTGA CATGACTTCT TTCATTTCAA TCCCTAATCT	60
15	TTCGGCAATT GTTCTAGCG CAATACCTTT TTGTGCATCT GAATGCGTTA TTTCTATATT	120
	TCCTCTCGAA GATGATGATA TAGCTAAATT CGGAGAKTCA GCTAAAATTT TACTAGCTTT	180
	GTCAATTTTT TCTAAATTTT CATCAAATGC TAATATTTTC ATAATTAATT CACCAGGTAT	240
20	GTTTTCAATA GCATCATAAT TATCAACAAC TyTCAACGTA CCATTATCTA TGGCTCTTTG	300
	AATACCATTT TTAATACGCT CAACGTTTGC ATGTTGACCT GCACGCTCAG CAATATCTAT	360
	GTAAATGTCT AAATCTCTTT GTGGATCTTC AGTATAAATC GCACGACTCG TGTATACTTG	420
25	ATAATAAATA CCTGCATCTT TTA AACATT TGTAATTTTG TGTACTAACG ATTTATTAAG	480
	GTGTGAAGTG CTCATTACAT TGAAAGTTTC ATCACGTA CTAGCACCAT TCAAACAAAT	540
30	ATATGGTACT GTTAAATCTG TGTCAGCAAC TGGTGCTTGk GCTTCATAAA ATGCTCGACC	600
	TGTCGCGATA ACAACCGTTA TCCCTTGTTT TTGAGCGTAT TTAATCGCAT CAATATTAGG	660
	TTGAGAAATT TCATGTGCTG CATTAAAGTAG CGTGCCATCC ATATCAGTGG CTATTAGTTT	720
35	TATCATTATG TnACCTCGTT TCGTAAATnT AAAATCTTGT TCTTAAATAA GrATATATAC	780
	TCAGCGCACA TACTTTtCTA TTAmCATTTA TATkGTCATT aATTTATCAT ATAATGTAAT	840
	TCTaACAAAT nTTAAcTAGT ATGTACTATC GTCTAATTGG TGGATTTCTT ATTGGCTCTT	900
40	AAgTTTTTAA AAAATGTTGT TAATAATGTG CTACATGCTT CTTTAAGTAC ACCTTTATCA	960
	ACAATTGCAC GATGATTAAA ATTAGATTGT TGCAATAAAT TCATTAAACT GCcACTACAA	1020
	CCACCTTTAG GATCATCTGC GCCATAGACG ACTCTTGGA TGGGACTCAT TACAATTGTT	1080
45	CCTGCGCACA TGACACATGG TTCTAAGGTT ACATATAATG TGCAACCTTC TAAACGCCAA	1140
	CTACCTAACA CTTTGGCTGC ACGTTCAATT GCAATATGTT CAGCATGCGC CGTTGGTTGT	1200
50	TGTAGTGT TTCTTaAATT ATGTGCTCTA GCGATAACTT CATCATCTTT AGTGATGATA	1260
	GCACCTATAG GTACTTCGCC TAGTTGAGCT GCTTTTTTAG CTCTTCAAT CGCTAATGTC	1320
	ATAAAATATA TATCATTTGT CATTATGTC CAGATACCTC ACTTATGGTA CAATACTCAA	1380

	CTATTGGCGT AGGTAAATCT TCACTTGACAC ACAAATTAAG TCAAACCTTTA GATTTTATG	1500
	AAGAAAAAGA AATCATCACA GAAAATCCAT TTTTATCAGA CTTTATGAA GATATCTCTA	1560
5	AATGGAGTTT TCAAAC TGAA ATGTTCTTTT TATGCAATAG ATATAAGCAA TTTCAAGATG	1620
	TAAACAACCT AAATCAAGGT GTAGTTAGTG ATTATCATAT ACATAAAAAAT AAGATATTTG	1680
	CTAAAAATAC TTTGAGTTCT GTTGAATTTC AGAAATTCAG TAAAAATTAT GATATTTTAA	1740
10	CTGAAGATAT GATTATGCCG AATATGATTA TCTTTT TAGA TGCAGACCTT GATGTGTTAA	1800
	AATCTAGAAT TGCTAAACGT AACCGTAGTT TTGAGCATCA AATAGAAgTG AtAcTGTAaG	1860
15	TTAAAAAAG ATTATCGTGA GTATTATGAG TCCTTACAAA GTAATGGTTC AAATGTAGTT	1920
	TTAATCGATA CnACTTCTAT TGATTTTCTT AAAAATGAAC AAGATTACGA AGATATATTA	1980
	CATATTATAT TACCTATGAT AGGAGATATT ACCAATGAAT AATTACGGTA TTCCACAAAA	2040
20	TGCCATTATA ACCATTGCAG GTACAGTTGG TGTGGAAAA TCAACACTAA sGCAAGCACT	2100
	TGCAGATAAA TTAAACCTTTA AAACGTCTTT TGAAATGTC GAACATAATC CATATTTAGA	2160
	TAAATTTTAC AGCGATTTTG AACGATGGAG TTTCCATTG CAAATTTACT TCTTAGCTGA	2220
25	ACGTTTTTAAA GAACAAAAGC GTATGTTTGA ATATGGTGGT GGCTTTGTCC AAGATCGATC	2280
	AATTTATGAA GATGTTGATA TTTTGC AAA AATGCATGAA GAAGAAGGCA CAATGAGTAA	2340
	AGAAGATTTC AAAACATATT CAGACTTATT TAATGCCATG GTCATGACAC CTTATTTTCC	2400
30	TAAACCTGAT GTAATGATTT ATTTAgAATG TAACTATGAT GAGGTCATTG ATCGTATTAT	2460
	TGAACGTGGT CGCGAAATGG AAATTAATAC AGACCCTGAA TACTGGAAAA AGCTATTTAA	2520
35	ACGCTATGAC GATTGGATTA ATAGCTTTAA TGCATGTCCA GTTGACGTA TCAATATTAA	2580
	TGAATATGAT ATCCATAAGG ACCCCGAATC TTTAAATCCT ATGATAAACA AAATTGCTCG	2640
	AATTATTCAA ACATATCGAC AAGTAGATAC ACGATAAAAG ACTAAAGACA TAGCGTATAT	2700
40	GTTTATATTC AATGTATATT CCATAGATAT TATCGATTAT TTTATCAATT CTATCGAATA	2760
	CATTAATTCA CATATACACT ATGTCTTTCT TTTTAATTTA AAGCTTCTAA AATATCTGCC	2820
	GCACTATTTA AAATAATATC AGCTTCATGT AATTCTTCTT TTGTTGCAAT ACCTGTTAAT	2880
45	ACACCTATTG CCATACCTAA ATTTGCATTA CTGCTGTCT TCATATCATT AGCAGTGTCT	2940
	CCTACTATAG CTACTTTCTG AGGATCTACA TTATATTGCT CAAATAAAGG CGATAATACT	3000
	TTAGGATTTG GCTTCTCATA GGCATCCGCT TCGGTAGAAA TGATCAAATC GAACAACGAG	3060
50	GTAGCATTGG TATGTGCTAA AAATTGTTCT ACACCTTTT TAGTATCACT CGTAACAATA	3120
	CCAAGTTGAT AGCCTTTTGC TTTCAAATCG ATAAGTGCTT CTTTAACACC TTCTACCCAA	3180
55		

	GTATCTTGTC	CCGTCACATC	ATTAAATGCC	TGGATAATTT	GTTGTAAAGA	TCCTGAACCC	3300
	ATCACTGATT	TTGGATCAAT	AGATTCTTTA	ATGACACCGA	GTTGTCTTAA	AGCAGCTTCT	3360
5	TTATTATGTA	CTGGGAAAGT	CTCAAGCAAT	GATTGTACAA	ATCGTACCCC	TATTTTTTCC	3420
	CAACTTCTAT	CAAATTCAAT	TAACGTACCA	TCTTTATCAA	ATAATATCCA	TTCCATTGaT	3480
10	ATCAATACTC	CTATTTATTT	ATTTTCGTATT	ATGCTGATTG	TATGATATTC	GTTATCCCCCT	3540
	GAAAATGAAC	TCGTAGTATT	GTTCTATTTA	AATATTGaAT	TAAATATAAT	AATAAGTGAA	3600
	ATCCCCTTCA	ATACTTAACA	ATAAACATTG	TAAACTTAAT	TTATTACCAT	GCTTCGCTTC	3660
15	ATTGAAAGGG	ATTTTAGTCA	TGATTAACTT	TGTCATATTG	TTTTCATGAT	TATATTCAAT	3720
	TTTTATTAAT	ATTTTGGTAC	AACGACTCTC	CAACCATTTT	TATCTTCTAA	AGTACCATTT	3780
	TGAATACCAG	TATAGACGTC	GTATAATTTT	TGAGTAATTT	CACCAGTCTC	ATTATTATTA	3840
20	ATAACGATTT	CACGATCTTC	GTATCTCAAT	GTACCCACAG	GTGAAATAAC	TGCTGCAGTA	3900
	CCACTACCAA	ATACTTCTGT	TAACTCACCT	TTATCATATG	ATTCGAATAA	TTCATCGATT	3960
	GAAACGCGGC	GCTCTTCGAC	TTCATATCCT	AAGTTTTTAG	CTAATTCGAT	AATAGATTTA	4020
25	CGTGTAATAC	CAGGTAAAAT	ACTGCCATTG	AACTCTGGTG	TAATTACTTT	GCCATTTTCA	4080
	ACGAAGAAAA	TGTTCATGCT	ACCAACTTCT	TCGATATATT	TCTGTTCAAC	ACCATCAAGC	4140
30	CATAATACTT	GGTCATAACC	TAATTTATTT	GCATTAGTTT	GTGCTAATAA	ACTTGcCGCA	4200
	TAGTTACCTG	CAACTTTTGC	AAAGCCTACA	CGCCACGaA	CAGCACGCAC	ATATTCACT	4260
	TCTACATAGA	TTTtagTTGG	TTTAAAGTT	TCACCACCAT	AATATGCACC	TGAAGGAGAT	4320
35	AAAATAATTA	ATAATTTATA	CTGATGTGAT	GCACCAACGC	CAAGTGCCCC	TTCTGTTGCA	4380
	AAAACAAATG	GACGAATATA	TAATGATTGA	CCTTCCCCCT	CAGGAATCCA	ATCTCTTTCA	4440
	ATATeAACTA	ATTGTTTTAG	CCCCTCTAAC	AATTCTGCTT	CGTCTACTTG	AGGCATTTCT	4500
40	AATCGTGCTA	ACGAGTTATT	AAGACGCTTA	AAATTTTCTT	CAGGACGGAA	AAGTGCAACT	4560
	TCCCCATCTC	TTTTATATGC	TTTTAATCCT	TCGAATACCG	ATTGACCATA	ATGAACACCT	4620
	TGTGCAGCAG	GTGAAATTTG	AATAGGACCA	TAAGGTACTA	TCTTCAAATC	ATGCCATCCT	4680
45	TTATCTGCAT	CATAATCATA	ACTCAACATA	TAATCAGTAA	AATATTTACC	AAAACCTAGT	4740
	TGAGATGTAT	TTGGTTTTTG	TTTAATGTT	TCTCGTCGTT	CAACTTTAAC	TGCTTGTGAC	4800
50	ATGGTGATTG	CCTCCTAATA	ATATTGTATA	AGAATTTGTT	TAACCTAAAT	TATAACAATC	4860
	CaTATTTTGC	TGTTCAACAA	ATTTTCTAAA	AATTCAAAAT	TAATTAACAG	ATTTCTAGAA	4920
	AGACTATATC	TTTtagTATA	AACGTATTAA	TTTCACAGAG	ACAAGTAATC	TGTGTTTTAC	4980

55

TAAGTATAAT GAATAATATT AGAATTCATG CACTAGTTTA TTAAAATAAA GAGTAATTTA 5100  
 AAATATCATT CCGTGTATTA AAGTGAATGG AAATGATTAG TTATTATTTT TAACAGTATC 5160  
 5 TTTTGTTC AATAGCTTCTA ACATTAATTT AGTCATGCTC GCTAAATCAT ATTTAGGATC 5220  
 AAATCCCAT TCGCCACGTG AACAGCTTGT ATCAATAGAA TCCGGCCAAC TATTAGCGAT 5280  
 ACCTTGTCTA ATAGGATCAA CATCGTAATC TAATGTAAAA TTGGGATAGT ATTCTTGAAT 5340  
 10 TGCTTCTTTT ACCATCTCTG GATCAAAACT CATTGCGCTC AAATTATAAC CATTCTAGT 5400  
 TTCTAATTTA GCGTCGTCTG CTTCCATAAG TTTAATAATT GCTTCAATTG CATCATCCAT 5460  
 ATACATCATA TCCATATACG TGCCTTTATC TATGAAGCTT GTATAATGAC CCTCTCTTAC 5520  
 15 TGCTTTGAAG TATATTTCAA CAGCATAGTC TGTAGTACCG CCACCTGGCT CTTTAACATG 5580  
 CGAGATTAAA CCTGGGAATC TAACACTTCT TGTATCTACA CCAAACGTT TGAAATAGTA 5640  
 20 TTGACACAAT AATTCTCCAG CTACTTTTATT TACACCATAC ATTGTCGTAG GTTGCTGAAT 5700  
 CGTTACTTGT GCGGTATTAA CTTTAGGAGT TGAGTCTCCA AATGCACCAA TTGAACTTGG 5760  
 TGTGAAAAG TGCAAATTAT AAGTTCTTGC AGCTTCTAAT GCATTCATTA ATCCACCCAT 5820  
 25 ATTTAAATCC CAAGCTAGAA TTGGATTTT CTCAGCAGTT GCTGATAATA ATGCTGCCAT 5880  
 ATGCATTAGA CTATCCGCTT CAAAGTCCCT AACTAACTCA AACATACGGT CACGATCTGT 5940  
 TACGTCTAAG ATTTCAAATG GTCCATTTTG TACAGGTGAG TCTGCTTCAG GTTCCCTAAT 6000  
 30 ATCTGTAGCA AGAACATTAT CTGTCCCAT AATTTCTCTG CACTTAACAA CTAATTCTGT 6060  
 ACCAATTTGT CCTAATGCAC CAGTAATCAT AATTTTTTTC ATAGAAATAT CTCCTTTGtC 6120  
 TCTTCTATAT AGCTATAGTC CATCACAAGC GgAcATAATA TTCATTTTCA TAATAATTAT 6180  
 35 AATATAAAG CGCTTTCTTG TATATATGAC ATGTACATGT TGCTGATATk TCTGTAAATG 6240  
 GAAATTCTAG TTGTATTAAT TGATTTTAGT AATTTATAGC GTTTATTATT GCTAATTACT 6300  
 40 GtCAAATTAA ATTTTTTATC CCTCAACTCT TAAACTCTGG ATATCTTTCA TTATATTAGC 6360  
 TTTTTTATAA CCATGGATAT CATGTAAAGC CTTATAAGCn TTAAATAATG TTTCATACCT 6420  
 TTGTACTTnT TCCGCTTCTG GATT 6444

45 (2) INFORMATION FOR SEQ ID NO: 234:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4721 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 50 (D) TOPOLOGY: linear

55

	GCCCATGAGA CAATTTTACT TGCTTTTCCC ATTGGTTATC ACGTTCTTAT TACATAGATT	60
	TAAACCGAGA AATATTATTC AAACGCTATT TATTGTATCG TTGATTTCTT TAGGACTTAT	120
5	GATAGTGATT CATTTCATCA CTGGAGATAA TTCACGTGTG TATTTTGGGA CAGATACACG	180
	ACTGCAAACT TTATTGCTTG GTTGATATATT AGCATTTATT TGGCCTCCGT TTGCTTTGAA	240
	AAAAGATATT TCTAAAAAGA TTGTCGTATC ATTAGATATT ATAGGGATAT CTGGTTTTCG	300
10	GGTTCTAATG ACTTTGTTCT TTATAGTTGG AGACCAAGAT CAATGGATCT ATAATGGAGG	360
	ATTTTACATT ATATCATTTG CAACTTTATT CATTATTGCA ATTGCGGTAC ATCCTTCTAG	420
15	TTTATTTGCT AAATTTTAA GTATGAAACC TTTACTAATT ATAGGTAAAC GATCATATAG	480
	CTTATACTTA TGGCATTATC CTATCATTGT TTTGTGAAC AGTTATTACG TACAAGGACA	540
	AATACCGTA TACGTTTATA TTATAGAAAT TTTGTTAACA GCGTTAATGG CTGAAATTC	600
20	GTATCGCTTT ATTGAAACAC CTATACGTAA AAAAGGATTT AAAGCTTTTG CATTTTACC	660
	TAAAAAGAAG GGGCAATTTG CTAGAACAGT GTTAGTTATC CTATTATTGG TTCCGCTAT	720
	CGTTGTGCTC AGTGGACAGT TTGATGCACT TGGCAAACAA CATGAAGCCG AGAAGAAAGA	780
25	GAAGAAGACG GAATTTAAAA CAACGAAGAA AAAAGTCGTT AAAAAAGATA AGCAAGAGGA	840
	TAAGCAGACA GCGAATAGCA AAGAGGATAT TAAAAAGTCA TCACCACTAT TAATTGGTGA	900
	CTCGGTCATG GTGGATATTG GTAATGTCTT TACTAAGAAA ATACCAAATG CACAAATTGA	960
30	TGGTAAAGTT GGACGGCAAC TCGTTGATGC TACACCAATT GTGAAATCGC AATATAAAGA	1020
	CTATGCTAAA AAAGGTCAAA AAGTTGTAGT AGAGCTTGGT ACAAATGGGG CATTTACGAA	1080
35	AGATCAATTA AATGAACAT TGGATAGTTT TGGAAAAGCA GACATATATT TAGTTTCTAT	1140
	TAGAGTACCT AGAGATTATG AAGGTAGAAT AAATAAATTA ATTTATGAGG CAGCTGAAAA	1200
	GCGCTCTAAT GTACATCTAG TCGATTGGTA TAAAGCTTCT GCAGGTCATC CGGAATACTT	1260
40	TGCATATGAC GGTATTCACT TAGAATATGC AGGTAGTAAA GCGCTGACTG ATTTGATTGT	1320
	AAAAACGATG GAAACACATG CTACAAATAA GAAATAATTT GATGCACTAA ACTTTTGAAA	1380
	TATTACATTA CTTCTGATAT TTATTATCAA AAATGATGTA TTTCATTAAA AGTTTAGTGC	1440
45	TTTTTTATTT TCAAATCCCA TAGTAACGGT GCAGAAAAAG TGTGTAAAC ATTCTAATTG	1500
	GTATATTACA TTCAATGAAG CTTTATTAGG AACAGATTAC ATTATGATAA CAAAGCCCGC	1560
	AAGACACCTA ATCTCTGTTA TAGTTTGTTC TGTCGCAAAA CTATAAAAGT TATAATTGTT	1620
50	TGCATACTAA AAAAAATAAA AATATAAAAT TAAAAATAAT TGAGTCGCTA ATGACTATAT	1680
	GTATAGAGTG TTTTGATTAT TGGGAGGATA TTAAATTATG AAAAAAATCG TTACAGCTAC	1740
55		

	ACAAAATAAC AATGGATATA ATTCTAATGA CGCTCAATCA TACAGCTATA CGTATACAAT	1860
	TGATGCACAA GGTAATTATC ATTACACTTG GACAGGAAAT TGGAATCCAA GTCAATTAAAC	1920
5	GCAAAACAAC ACATACTACT ACAACAATA CAATACTTAT AGTTATAACA ATGCATCTTA	1980
	CAATAACTAC TATAATCATT CATATCAATA CAATAACTAT ACAAACAATA GTCAAACAGC	2040
	AACAAATAAC TATTATACTG GTGGTTCAGG TGCAAGTTAT AGCACAACAA GTAATAATGT	2100
10	TCATGTGACT ACAACTGCAG CGCCATCTTC AAATGGTCGT TCAATTTCTA ATGGTTATGC	2160
	ATCAGGAAGT AACTTATATA CTTCAGGACA ATGTACTTAT TATGTATTTG ATCGTGTGG	2220
	TGGGAAAATT GGTTCACAT GGGGTAACGC AAGTAATTGG GctAACGCAG CTGCATCATC	2280
15	TGGCTATACA GTGAACAATA CACCAAAAGT TGGTGCTATC ATGCAAACAA CACAAGGCTA	2340
	TTACGGTCAT GTTGCTTACG TTGAAGGCGT TAACAGCAAC GGTTCTGTTC GTGTTTCAGA	2400
	AATGAACTAT GGACATGGTG CTGGTGTGGT TACGTCTCGT ACAATTTTCAG CAAACCAAGC	2460
20	AGGTTTCATAT AATTTTCTTC ATTAATCAAA TGTAATCAA ATGACGTCAA TATTCTCTAA	2520
	CATGAGAGTA TTGGCGTTTT TGTTTTATAT AAATATAAAT GAGAGCGGTT TATTCCTGA	2580
25	TCTTTAGGGA ACTAAGTAAT AAAGTGATAA TTTATACTAT GTCAGTATGA TTGAGAGTGA	2640
	TTCAATTTAG ATGAAAACCA TGA AAAAATA TATTAAAACA GCATTTTTTT GTAGTATGTA	2700
	TTGGTTAATT GTTCAACTAA ATATAGCAAA TTTAGGTACA AGAATTCCTG ATAAGTATTT	2760
30	TCGTCAGAAG TACATAATAT TTAAATCATT TAACTTTGAG AAGCATGGAA AATTTTGGAA	2820
	CAATGGTTT TACGTAAGAA AATGGAAACA TAAGATTTTA GATGGTCATC AGCTTAATCA	2880
	AAATATATAT GATCAGCGTC ATTTAATGAC AATCAATACT GATGAAATTG AAAAAATGAT	2940
35	TATAGAGACA AAGAGGGCAG AGTTGATTCA TTGGATATCG ATACTTCCAG TCATCATATT	3000
	CAATAAAGGC CCTCGTTTAG TAAAGTATAT AAATATTTTC TATGCAATGA TAGCTAATGT	3060
	TCCAATCATT ATTGTGCAAC GCTATAATCG ACCGAGATTA ACGCAGTTAC TACGCATATT	3120
40	AAAACGAAGA GGTGAACGTC ATGACTAAAC ATATCATCGT TATTGGTGGT GGCTTAGGTG	3180
	GGATTTCTGC AGCAATTCGA ATGGCACAAA GTGGCTATTG GGTCTCATT TATGAACAAA	3240
45	ATAATCATAT AGGAGGCAAA GTGAATCGTC ATGAATCAGA TGGCTTTGGC TTTGATTTAG	3300
	GTCCATCTAT TTTAACGATG CCTTATATTT TTGAAAATT ATTCGAATAT AGCAAGAAGC	3360
	AAATGTCAGA CTACGTTACA ATCAAGCGAT TGCCACATCA ATGGCGTAgC TTTTTTCCAG	3420
50	ATGGAACGAC TATCGATTTG TATGAAGGTA TTAAAGAAAC AGGTCAGCAT AATGCGATAT	3480
	TGTCGAAACA GGATATAGAG GAACTGCAAA ATTATTTGAA TTATACAAGA CGAATCGATC	3540

55



TTCATGGGCC ATTAAATGCT CTTATTAATT ATGATTATGT ACATACTATG CAACAGGCCA 3660  
 TAGACAAGCG TATCTCGAAT CCATACTTGC GACAAATGTT AGGCTATTTT ATCAAATATG 3720  
 5 TAGGTTCTTC ATCATACGAT GCGCCAgCTG TATTATCTAT GTTATTCCAT ATGCAACAAG 3780  
 AGCAAGGCCT TTGGTATGTA GAAGGTGGAA TCCATCATTT AGCCAATGCC TTGAAAAGc 3840  
 tAGCGCGTGA AGAAGGTGTC ACAATTCATA CAGGTGCACG TGTGGACAAT ATTAAAACAT 3900  
 10 ATCAAAGACG TGTGACGGGT GTCAGATTAG ATACAGGTGA GTTTGTAAAG GCAGATTATA 3960  
 TTATTTCAAA TATGGAAGTC ATACCTACTT ATAAATATTT AATTCACCTT GATACTCAAC 4020  
 GATTAAACAA ATTAGAGAGG GAATTTGAGC CGGCAAGCTC AGGATATGTG ATGCATTTAG 4080  
 15 GTGTGCTTG CCAATACCCG CAATTAGCAC ATCATAATTT CTTTTTTACG GAAAATGCTT 4140  
 ATCTCAATTA TcAACAAAGTT TTTcATGAAA AGGTATTGCC AGATGATCCG ACCATTTATC 4200  
 TAGTAAATAC GAATAAAACT GATCACACAC AAGCGCCAGT AGGTTATGAA AATATCAAAG 4260  
 TCTTACCACA TATTCCATAT ATTCAAGATC AGCCTTTTAC CACTGAAGAT TATGCGAAGT 4320  
 TTAGGGATAA AATTTTGGAT AAATTAGAAA AAATGGGACT TACTGATTTA AGAAAACACA 4380  
 25 TTATTTATGA AGATGTTTGG ACACCGGAGg ATATTGAAAA AAATTATCGT TCTAATCGTG 4440  
 GTGCAATATA TGGTGTGTA GCAGATAAAA AGAAAAACAA AGGATTTAAA TTTCTAAAG 4500  
 AAAGTCAGTA TTTTGAAAAC TTGTACTTTG TAGGTGGATC AGTAAATCCT GGTGGTGGA 4560  
 30 TGCCAATGGT TACATTAAGT GGGCAACAAG TCGCAGcAAg ATAAACGCGC GAGAAGCGAA 4620  
 GAATAGGAAG TGATATCTAT GAAATGGTTA TCACGAATAT TAACAGTAAT AGTGACCATG 4680  
 TCTATGGCGT GTGGTGcATT GATATTTaAT CgTAGACATC A 4721

## (2) INFORMATION FOR SEQ ID NO: 235:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3516 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

TATTCGTGCG CAATGGGCTA AATTAGGTCT AGGTTTAGAT TATAGTAGAG AACGTTTTAC 60  
 TTTAGATGAA GGTTTAAGTA AAGCAGTTAA AAAAGTTTTT GTTGATTAT ACAATAAAGG 120  
 50 AATTATTTAT CGTGGCGAAC GTATTATAAA TtGGGATCCn AAAGCACGTA CAGCTTTATC 180  
 TGaTATTGAA GTAATACATG AAGATGTTCA AGGTGCGTTT TATCATTTTA AATATCCTTA 240

## EP 0 786 519 A2

	TACAGCGATT GTTGTTAACC CTAATGACGA ACGATACAAA GATGTAATCG GTAAAACTGT	360
	TATATTACCA ATCGTAGGAC GCGAACTGCC TATTTTAGCA GATGAGTATG TTGATATAGA	420
5	CTTCGGTTCT GGTGCTATGA AAGTGACACC AGCACATGAC CCTAATGATT TTGAAATTGG	480
	TCAAAGACAT CAATTAGAAA ATATTATCGT TATGGATGAA AATGGTAAAA TGAACGACAA	540
	AGCGGGTAAA TATGAAGGTA TGGACCGTTT TGATTGTCGT AAACAGCTAG TTAAAGATTT	600
10	AAAAGAACAA GATTTAGTTA TCAAGATTGA AGATCATGTT CATTCTGTAG GTCATTGAGA	660
	ACGATCTGGC GCTGTTGTTG AACCATATTT ATCAACACAA TGGTTTGTGC GCATGGAAGA	720
	CTTAGCGAAA CGTTCATTAG ATAACCAAAA AACAGATGAT CGTATTGATT TTTATCCGCA	780
15	ACGTTTCGAA CATACTTTA ACCAATGGAT GGAAAATATT AGAGATTGGA CGATTTCAAG	840
	ACAATTATGG TGGGGTCATC AAATTCCGGC TTGGTATCAT AAAGAAACAG GCGAAATATA	900
20	TGTTGGAGAA GAAGCGCCAA CTGATATTGA AAATTGGCAA CAAGATGAAG ATGTATTAGA	960
	TACGTGGTTC TCaAGTGCTT TATGGCCTTT CTCyACGTTA GGTGGCCTG ATTTAGAAAG	1020
	TGAAGACTTT AAACGATACT ACCCAACAAA TGCCTTAGTT ACAGGTTACG ATATTATCTT	1080
25	TTTCTGGGTA GCACGCATGA TATTCCAAGG CTTAGAATTT ACAGATCGTC GTCCATTTAA	1140
	TGATGTATTA TTACACGGTT TAGTTCGTGC TGAAGACGGG CGTAAGATGA GTAAATCATT	1200
	AGGTAATGGT GTGGATCCAA TGGATGTTAT TGACGAATAC GGTGCTGATA GCTTGCCTTA	1260
30	CTTCTTAGCA ACAGGTTTCAT CTCCAGGACA TGATTTAAGA TACTCAACTG AAAAAAGTTGA	1320
	GTCAGTGTGG AACTTTATCA ATAAAATCTG GAATGGGGCA CGTTTCAGTT TAATGAATAT	1380
	CGGTGAAGAC TTTAAAGTTG AAGATATCGA TTTAAGTGGT AACTTATCAT TAGCAGATAA	1440
35	ATGGATTCTA ACACGTTTAA ATGAAACGAT TGCAACAGTT ACTGATTTAA GTGACAAATA	1500
	TGAATTCGGC GAAGTTGGAC GTGCATTATA TAATTTCATT TGGGATGATT TCTGTGATTG	1560
40	GTACATTGAA ATGAGTAAAA TTCCAATGAA TAGTAATGAT GAAGAACAAA AACAAGTTAC	1620
	ACGTTTCAGTA TTGAGTTATA CTTTAGACAA TATTATGAGA ATGCTACATC CATTGATGCC	1680
	ATTTGTAACA GAGAAAATAT GGCAAAGTTT ACCACATGAA GGTGACACAA TTGTTAAAGC	1740
45	TTTCATGGCCA GAAGTGCCTG AATCATTGAT TTTTGAAGAA AGTAAACAAA CAATGCAACA	1800
	ACTTGTTGAA ATCATTAAAT CTGTAAGACA ATCACGTGTA GAAGTAAATA CGCCATTGTC	1860
	TAAAGAAATA CCTATTTTAA TTCAAGCTAA AGATAAGAA ATTGAAACAA CTTTATCACA	1920
50	AAACAAAGAT TATTTAATCA AATTCTGTAA TCCTAGTACC TTAAATATTA gCtGACGTGG	1980
	AAAwTCCTGA GAAAGCAATG ACaTCAGTTG TAATTGCAGG TAAAGTGGTA TTACCATTAG	2040

55

AAAGCGAATT AGATAGAGTA GATAAAAAGC TCTCTAATGA AAACTTTGTA AGTAAAGCAC 2160  
 CTGAAAAGGT TATAAATGAA GAAAAACGTA AAAAAACAAGA TTATCAAGAA AAATATGATG 2220  
 5 GTGTGAAGGC AAGAATTGAA CAATTAAAAG CATAGGAGTT AGTAACAATG AATTACCTAG 2280  
 AGAGCTTGTA TTGGATACAC GAAAGAACTA AATTTGGCAT CAAACCAGGT GTTAAACGTA 2340  
 TGGAATGGAT GCTAGCACAA TTTAATAATC CTCAAATAA CATTAGGGT ATTCATGTAG 2400  
 10 GTGGCACAAA TGGTAAAGGC TCTACAGTTG CTTACCTTAG AACAGCTTTA GTTGAAAATG 2460  
 GTTATGAAGT AGGTACATTT ACGTCGCCGT TTATTGAAAC ATTTAATGAA CGAATTAGTC 2520  
 TAAATGGTGT GCCAATATCA AATGACGCTA TTGTAGAATT AGTATCACGT ATTAAACCAG 2580  
 TAAGTGAAAT GATGGAACGT GAAACAGATT TAGGTGTTGC AACTGAATTC GAAATAATCA 2640  
 CAGCGATGAT GTTTTATAT TTTGGTGAAA TACATCCTGT TGATTTTGTC ATTGTTGAGG 2700  
 20 CTGGATTGGG TATAAAGAAC GATTCGACAA ATGTCTTTAC ACCGGTTTTA TCAATCTTAA 2760  
 CTAGTATCGG TCTAGACCAT ACAGATATTT TAGGTGGTAC TTATCTAGAT ATTGCTAGGG 2820  
 ATAAAGGCGC GATTATAAAG CCTAACGTTT CAGTGATATA TGCTGTTAAA AATGAAGATG 2880  
 25 CATTAAAATA TGTTCTGTAA CGCGCAATTG AACACATGC AAAGCCAATT GAATTAGATA 2940  
 GAGAAATTGT TGTTGTATCG CAAATGATG AATTTACTTA CCGTTATAAA GATTATGAAT 3000  
 TAGAAACAAT CATTTTAAGC ATGTTAGGTG AACATCAGAA ACAAATGCT GCATTAGCCA 3060  
 30 TAACAGCTCT TATTGAATTA AATGAACAAG GATTAATTGA ATTAGATTTT AATAAGATGA 3120  
 TAGACGTAT TGAATCAGTT CGTTGGACTG GACGTATTGA GCAGGTGCAT GACAAACCTT 3180  
 TAATCATTTT GGATGGCGCA CATAATTCAG AGAGTATAGA TGCTCTAATT GATACAATTA 3240  
 35 AACAGTACCA TGATAAGAA AAAGTAGATA TTTGTTCTC AGCAATAAAC GGAAAACCGA 3300  
 TTAACGAGAT GGTCAAACAT TTAAGTTTAA TTGCGCATAC GTTTTATGCA ACTGAATTG 3360  
 ATTTTCCGAA AGCGTTACGC AAAGAAGAAA TTGTAGGTAG TATTGAAAAT GATGAAATAC 3420  
 40 AATTAGTAGA TGAATACGTT GAATTTATAA AAAATTATCA AGGTGATACA TTAGTAATTA 3480  
 CCGGTAGTCT GTATTCATA AGTGAAGTTA AATCAA 3516

45 (2) INFORMATION FOR SEQ ID NO: 236:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7481 base pairs  
 (B) TYPE: nucleic acid  
 50 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

	TGAGTGATAG AATCAAAAAA GCCATCTCAA AAATTAATCA AGCAAACAAC ATTCCAAACA	60
	ATGsTCGCAA ATCACC AATG TATCACTCTC CAATTACGTA ACTATGATTT AATTTAAGCA	120
5	TAGTTATTGA GGT TTTGTGA TATATAGTAT AAAATTAATG AGAATTAAAT TTAATAATGT	180
	AAAATTCATm TTCgGGGTCG GGTGTAATTC CCAACCGGCA GTAAATAAAG CCTGCGACCT	240
10	GCTAGTATGT ATCATATTAG TGGCTGATCT AGTGAGATTC TAGAGCCGAC AGTATAGTCT	300
	GGATGGGAGA AGATGGAGGT TTTTGTGTGT GCAATAATCC TCCTATTCTT ACGAGATGAA	360
	TGGAAGGAGA AAATTGAATA TGCAACAAA TAAACGTCTT ATCACAATAA GTATGTTGAG	420
15	CGCGATTGCG TTTGTGTTAA CTTTTATCAA GTTTCCTATA CCATTTTTGC CACCATACTT	480
	AACTTTAGAT TTTAGTGATG TACCGTCACT ACTAGCTACA TTTACGTTTG GACCAGTTGC	540
	CGGTATCGTA GTTGCACTGG TTA AAAATTT ATTGAACTAC TTATTTAGTA TGGGCGATCC	600
20	AGTTGGACCA TTTGCTAACT TTTTAGCAGG CGCAAGTTTC TTATTAAGTG CTTACGCCAT	660
	CTATAAAAAT AAACGTTCAA CAAAATCTTT GATTACTGGA TTAATCATTG CAACAATCGT	720
	TATGACTATC GTGTTGAGTA TTTTGAAC TA TTTGTTCTA TTACCTTTGT ACGGTATGAT	780
25	ATTTAACTTA GCTGATATCG CAAATAATCT TAAAGTAATC ATTGTTTCAG GAATTATACC	840
	ATTCAATATT ATTAAAGGTA TCGTTATTTT TATTGTATTT ATTTTACTAT ATAGAAGGCT	900
30	TGCGAATTTTc TTGAAAAGAA TTTAATCAAA TTAAAGCAAA ATAATATACA CATAATAATA	960
	AAAAGCAGGT GACTATCAAT AAACGATAGC TTGCCTGCTT TTTCTATAGA ACATTTGTCT	1020
	AAAAAATCAA TTATTCaAAT TTTAAAGCGT CCCCATCAAA TGATTGCTCT GCAATTTTAA	1080
35	TAGAATCTGt AGGGCATCCa TCAATTGCAT CTTCCATATC TTCATATAAT TCCTCAGGTA	1140
	CTTCTGCAGT ACCTTGGTTA TCGTCAAGGA TTACGAAAGC AATACCTTCG TCGTCGTAAT	1200
	CATATATATC TGGCGCTGCT GCACCGcATG CACCACATGC AATACAAGTA TCCATATCAA	1260
40	CGATTGTATA TTTTGCCAAT GTCTTCGCCT CCTTTGATAA AAATGCTAAA ATAGTAATGT	1320
	GACTAAAATT TTAGACAGCA TCATTTTTAT TTTCAAATTA TCCGTTTTAC AGAGTGAGGG	1380
	TTAAATTTGC AACACATTAT AAAAACAGCA TTACAACAAA CATTTAAC TA AAAACAAAT	1440
45	AAAAGTATTT ACAATATCTT AGTTGGTAAG AAATCTCACC AAACCTTTTT TGACGCTTGT	1500
	AGTCAACAAC AGTTGTCATT ATATCACAGT TTACCACTAT TAAAATATCC GTCTTTTGAG	1560
	CTATTTtTAG AAAAAATCAA TGAATTTAAT GCTGAAATGG AAATCATGTT GCATCCTAGA	1620
50	TATACATTTG aAAGCATGGG GCAAACATTT CAAGCAATTC AACTATTAGT GCAAaCCATG	1680
	TCTAATACCA AACAACATGT TTTTCATTTT GTACCAATCT CTCAAAATAA TAAGATACAm	1740

55

## EP 0 786 519 A2

	AATGAACTAC	ATAACTTGTT	TAAAGCAATA	ACTTTAAAAG	GGCCATGTTA	CTTACATTAT	1860
	TATTTGCAAG	GCTATGATGA	ACCAATGTAT	ACGAGACAGC	AAGTTAGTTT	AATAGAAAAG	1920
5	CTATCTCAAC	AGCAATTGTT	TGAATACGAA	ATGAATAATT	TAGTGACAAT	GATGTTTGAA	1980
	TTAGAAAGTG	GAGAATATAC	TATTTTATCA	AAAATAATAA	TGAAACCTAC	ATTATTAAAT	2040
	CAAACTTATA	TTACTTATAC	AAAATTGCTT	GAACAATTCA	CGATGGAAGA	TATAGCGGCT	2100
10	CAACAACAAG	TTAAAATCAA	TACTATCGAA	GATCATGTAC	TTGAAATCTT	AATCAAAGGT	2160
	TACATGTCTA	ATTACGATGA	TTATGTTGAA	CTAGAAGATC	AACTCCAGTT	TTTGAATTTT	2220
15	TATCAACAGC	ATCGTGGCGA	ACGATTAAAA	TTTACAAAG	AACAATTTGA	CACGTTATCA	2280
	TATTTTCAAT	TAAAAGTATT	AATCGTTGGA	TTTGAAAGAG	GTGATCTGAA	TGTTGCATGA	2340
	TATTTTACGA	AACAAATTTG	GATTTCGAGAG	TTTTAAACCG	GGACAACAGG	AAATTATAGA	2400
20	AAGTATAATG	TCTCAACAAC	ACACTCTAGG	TATACTTCCA	ACTGGAAGTG	GAAAGAGTTT	2460
	GTGTTATCAA	ATACCTACGT	ATTTATCAGG	TAAGCCGACA	TTAATTATCT	CACCGTTAAT	2520
	ATCTTTAATG	GATGACCAAG	TTATGCAGTT	GAAAATAAAT	GGAGAAAAAC	GTGTAACATG	2580
25	TATTCACTCT	GGTATGGATG	AAATTGAGAA	AAAGCATAAT	ATTAAATGTT	TACGACATAG	2640
	CCGCTTCATC	TTTCTAAGTC	CAGAATTTCT	CCTGCAACCG	TCAAATTTTA	AATTAATATC	2700
	TATGATAGAC	TTTGGCATGA	TTGTTCTAGA	TGAAGCACAT	TGCCTATCTG	AATGGGGATA	2760
30	TGATTTTACA	CCACATTATG	CTCTAATAGG	AAAAGTAACA	AAGCATTTTA	AAGAAGCGGT	2820
	TGTCTTAGCA	TTGACAGCAA	CTGCACCACC	GCATTTACAA	GATGATTTGA	CGGAAATGTT	2880
35	AGCGATTCAA	TTCAATGTTA	TTAAAACCTAC	AATGAATCGC	CCAAATATAA	GCTTTAAGCA	2940
	TCTTAATTTT	CATGATGATG	AAGATAAAAT	TGAATGGTTG	CTGCCGTTTC	TACAACAGTC	3000
	GGGAGCAACG	ATTATTTATG	TCTCATCGAA	AAAGATGTGT	CTGAATTTAG	CGCAACTTAT	3060
40	TTATGATTCA	GGTTTTCTTA	CAGGTATTTA	TCATGGTGAT	ATGAATTATC	AAGAGCGACA	3120
	CACAGTTCAA	CAACAATTTT	TAAATAATGA	TATTCCGATT	ATAGTCGCAA	CGAGTGCTTT	3180
	TGGTATGGGA	ATTAATAAAA	AAGATATTCTG	CACAATCATT	CACTTTCATC	TTTCAACAAG	3240
45	TCCTTCTAAC	TACATTCAAG	AAATTGGCCG	TGCGGGTCGC	GATGGTGAAC	TAAGTCAGGC	3300
	AATTAGTTTA	TTCCAACCGG	ACGATAAATA	TATTTTAGAA	ACGTTATTAT	TTGCAGATAT	3360
	GATAACAGAA	GAAGATGTAC	AAAATTTCGA	AATAGGAGAA	TTTTTAGCTC	CCGATAAACA	3420
50	AGCCGTTTGT	ACAACGTTGc	AATCATTCTA	TAGTATCGGC	GCCTTGaAAC	AGATATTTAA	3480
	GCAATCATTT	AAACGAAAGC	AATTAGGATT	CTTTCGCATG	ATTGGCTATT	GCAAATTGGA	3540
55							

	ATGTTGTGAC AATGATTCTA ATATAACTGA TATCGCAATT TTAAATAAGA AGAAGGTAAT	3660
	TAGAAGTATT GGATTGTATG AAAAGTTGCA AAATTTATTT CTCAGATAGT ATTACTTTAC	3720
5	TAAAAGAAAA TTGACAAGCT ATAATTAGTG TATACACAAT TGAAAAATGA TTGAAATAAT	3780
	TTTGAAAAAT ATACATAAAC ATATGTCATG TGGGTATATT TTATGTAAAA TCATTGTAAT	3840
10	AGAATAGAAA GGAAGATGGC TATGTCTAAT AATTTTAAAG ATGACTTTGA AAAAAATCGT	3900
	CAATCGATAG ACACAAATTC ACATCAAGAC CATACGGAAG ATGTTGAAAA AGACCAATCA	3960
	GAATTAGAAC ATCAGGATAC AATAGAGAAT ACGGAGCAAC AGTTTCCGCC AAGAAATGCC	4020
15	CAAAGAAGAA AAAGACGCCG TGATTAGCA ACGAATCATA ATAAACAAGT TCACAATGAA	4080
	TCACAAACAT CTGAAGACAA TGTTCAAAAT GAGGCTGGCA CAATAGATGA TCGTCAAGTC	4140
	GAATCATCAC ACAGTACTGA AaGTCAAGAA CCTAGCCATC AAGACAGTAC ACCTCAACAT	4200
20	GAAGAGGAAT ATTATAATAA GAATGCTTTT GCAATGGATA AATCACATCC AGAACCAATC	4260
	GAAGACAATG ATAAACACGA TACTATTAAA AATGCAGAAA ATAACACTGA GCATTCAACA	4320
	GTTTCTGATA AGAGTGAAGC TGAACAATCT CAGCAACCTA AACCATATTT TACAACAGGT	4380
25	GCTAACCAAT CAGAAACATC AAAAAATGAA CATGATAATG ATTCTGTAAA ACAAGATCAA	4440
	GATGAACCTA AAGAACATCA TAATGGTAAA AAAGCAGCAG CTATTGGTGC TGGAACAGCA	4500
30	GGTGTTCAG GTGCAGCTGG TGCAATGGCT GCTTCTAAAG CTAAGAAACA TTCAAATGAC	4560
	GCTCAAAACA AAAGTAATTC TGGCAAGGCG AATAACTCGA CTGAGGATAA AGCGTCTCAA	4620
	GATAAGTCTA AAGATCATCA TAATGGCAAA AAAGGTGCAG CGATCGGTGC TGGAACAGCA	4680
35	GGTTTGGCTG GAGGCGCagC AAGTAAAAGT GCTTCTGCCG CTTCAAACC ACATGCCTCT	4740
	AATAATGCAA GCCAAAACCA TGATGAACAT GACAATCATG ACAGAGATAA AGAACGTAAA	4800
	AAAGGTGGCA TGGCCAAAGT ATTGTTACCA TTAATTGCAG CTGTACTAAT TATCGGTGCA	4860
40	TTAGCGATAT TTGGAGGCAT GGCATTAAAC AATCATAATA ATGGTACAAA AGAAAATAAA	4920
	ATCGCGAATA CAAATAAAAA TAATGCTGAT GAAAGTAAAG ACAAAGACAC ATCTAAAGAC	4980
	GCTTCTAAAG ATAAATCAAA ATCTACAGAC AGTGATAAAT CAAAAGAGGA TCAAGACAAA	5040
45	GCGACTAAAG ATGAATCTGA TAATGATCAA AACACGCTA ATCAAGCGAA CAATCAAGCA	5100
	CAAAATAATC AAAATCAACA ACAAGCTAAT CAAATCAAC AACAGCAACA ACAACGTCAA	5160
	GGTGGTGGCC AAAGACATAC AGTGAATGGT CAAGAAAACT TATACCGTAT CGCAATTCAA	5220
50	TACTACGGTT CAGGTTCAAC GGAAAATGTT GAAAAAATTA GACGTGCCAA TGGTTTAAGT	5280
	GGTAACAATA TTAGAAACGG TCAACAAATC GTTATTCCAT AATATAACTA TATAAATTGT	5340
55		

	TAAATTGCGC TTATAAGTAT GTAGCGGTTT TTTCATTTTT CAAAGTTTGT TATTTAACAA	5460
	GGTCTTGTCT CGAATATTGG CATATCAATT TAACTTTTTA AATAGTCATC AAAAAGATAA	5520
5	AACACCACAA TCAACAAATT TAACGAGGAA GAATAAAAAA TAAATCAACA TATTAAATTG	5580
	TAGTGTATT CAACTCCGTA GCTAACAATT CTCTATTAC ATTAAACAA TTGTCAAAAA	5640
	TATATCATAA ATCTTCAAGC ACAGACTTAG CGCATCAATC ACTGAACTGT TATAATAGTT	5700
10	TGGGATTAAA GGAGGCCGAA ACAATGCAAA AAGTTGAAAG TATCATAATT GGTGGAGGGC	5760
	CATGCGGATT AAGTGGGCT ATTGAACAAA AAAGAAAAGG TATTGATACC TTAATTATTG	5820
	AAAAGGGTAA TGTCGTTGAA TCAATCTACA ATTATCCTAC TCACCAAACA TTTTCTCAT	5880
15	CAAGTGATAA ATTAAGTATT GGGcGAgTAC CGTTTATCGT TGAAGAAAGT AAACCAAGAC	5940
	GTAATCAAGC GCTAGTTTAT TACCGAGAAG TTGTAAAACA TCATCAATTA AAAGTAAATG	6000
20	CATTTGAAGA AGTATTAAGT GTTAAAAAAA TGAATAATAA ATTTACTATT ACTACGACGA	6060
	AAGATGTTTA TGAATGTCGA TTTTAAACAA TCGCGACAGG CTATTATGGT CAGCATAATA	6120
	CATTAGAAGT TGAAGGTGCG GATTTACCTA AAGTGTTCCTA TTATTTTAAA GAGGCACATC	6180
25	CGTATTTTGA TCAAGATGTT GTAATTATCG GTGGTAAGAA TTCGGCTATC GATGCTGCTT	6240
	TGGAGTTGGA AAAAGCTGGT GCTAACGTGA CGGTTCTATA TCGTGGTGGA GATTATTCGC	6300
	CTTCAATTAA ACCGTGGATA CTTCCAAATT TCACAGCATT AGTAAATCAT GAAAAAATTG	6360
30	ACATGGAATT TAATGCTAAT GTTACCCAAA TAACTGAAGA TACTGTGACT TATGAAGTAA	6420
	ATGGTGAAAG TAAAACGATA CACAATGATT ATGTATTTGC GATGATTGGT TATCATCCCG	6480
	ATTATGAATT TTTAAATCT GTAGGCATTC AAATTAATAC AAATGAATTT GGAACAGCGC	6540
35	CTATGTATAA TAAAGAAACA TACGAAACAA ATATCGAAAA TTGCTATATT GCAGGTGTAA	6600
	TTGCTGCAGG GAACGATGCG AATACCATTT TTATTGAAAA TGGTAAATTC CACGGGGGCA	6660
40	TTATTGCTCA AAGCATGCTA GCTAAGAAAC AAACGCCCTT AGAATCATAA AAATAAAGGT	6720
	CTATGTACTA AATAACTTAG TTTTACAACG ACTGACATTC ATGATATGTC AAATGAGGTT	6780
	GATGACTATT GATTGTACAT AGACCTTTTT ATGTTACGTA TTCATTATAA TTCAAAATAT	6840
45	GATTTGATTT CAGCTTTATC TAAATTGTTG CTTAACGCGA CTAATAATTT TAATCTTGCT	6900
	TTTGGACCAT TCAAGCCGTT AGAAAAATA AAACCTTGTT GTGCGAGTTG GTAACCACCA	6960
	CCATCGTATG CGTAAGTTGG ACTCACAATA CCATTAAAGG AACGTGAAAC TAGCACAATA	7020
50	GGTATATTTA AAGATACTAA TTGTTGAATG CCTTCTAATG CGCTTGAGG TATGTTGCCT	7080
	TGTCCTAACG CTTCAATAAC CATACCATCG ATACCTTCTC GACTATAAAA ACTAAAAATG	7140

55

TCCAATGCTT GTTGGCGATA TGGCATATGA TGGAAATTGCA CACGATCCTT TGTCAATACA 7260  
 CCTAGCGGAC CATGATTTGG ACTTTGAAAT GTGTTTGTAT TAGACGTATG TGTTTTGGTA 7320  
 5 ACATTACGCG CCGTATGAAT TTCATCATTA AaTACAACCA TCACGCCTTT ATGACGGGCC 7380  
 TTTTCATCAG AGGCAACGCG AATAGCGGAA ATATAATTAT ATAATCCGTC AGAACCAATT 7440  
 TCATTAGACG AGCGCATTGG CCAAGAAntA ACAACAGGTT G 7481

10

(2) INFORMATION FOR SEQ ID NO: 237:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6346 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

ATGAATTATA CCTTCCACTA GAAGTGTCCG TATTTATGTG CCTGGTGGTA AAGCAAGTTA 60  
 TCCTTCAACG GTATTAATGA CAGCGACGTT AGCACAAGTA GCAGGTGTGG AAAATATTGT 120  
 25 CGTTGTGACA CCACCTCAAC CTAATGGTGT GTCTCAAGAA GTATTAGCAG CATGCTATAT 180  
 TACGCAAGTT GATCAAGTAT TTCAAGTTGG TGGTGCTCAA AGTATTGCTG CGTTAACTTA 240  
 TGGCACAGAA ACGATACCTA AAGTTGATAA GATTGTAGGT CCAGGCAATC AATTTGTTGC 300  
 30 ATATGCCAAG AAATATTTAT TTGGACAAGT AGGTATTGAC CAAATAGCTG GACCAACAGA 360  
 AATAGCACTT ATTATTGATG AAACAGCAGA TTTAGATGCC ATCGTATATG ATGTTTTTGC 420  
 ACAAGCAGAA CATGATGAAT TAGCACGTAC ATATGCCATT AGTGAAGATG CGCAAGTCCT 480  
 35 TAAAGATTTA GAATCACGTA TTGCTAAAGC ATTGCCTAAT GTGGACAGAT ACGACATTGT 540  
 TTCTAAAAGT ATCGCTAATC AACACTACCT TATCCATGCT AGTAATTTTG ATGAAGCATG 600  
 CCATGTCATG AATACAATCG CGCCTGAACA TGCCTCGATT CAAACAGTAA ATCCTCAACC 660  
 40 ATATATTGAA AAAGTGAAAT ATGTGGGTGC ATTGTTTATT GGACATTATT CGCCAGAGGT 720  
 CATAGGAGAT TACGTTGCAG GTCCAAGTCA TGTATTACCT ACAAATAGAA CAGCTAGATT 780  
 45 TACCAATGGG TTATCGGTCA ATGATTTCTT AACACGGAAC ACGGTCATCC ATTTATCAAA 840  
 AGATACGTTT GAACAAATTG CTGATTCAGC ACAACATATT GCTCATGTTG AAGCATTATA 900  
 CAATCACCAG CAGTCTATTT TAATACGTCA GTCTTAGGGG AGTGTAAATTG AAATGATTTA 960  
 50 TATTGATAAA AATGAAAGTC CAGTTACGCC GTTGATGAA AAAACAATGA CGTCTATTAT 1020  
 TAGTGCAACg CnATATAATT TATATCCTGA TGCAGCATAT GAACAATTCA AGGAAGCTTA 1080

55



## EP 0 786 519 A2

	GATTCAAAAG TTAATGCTGA TCATGCCAGA AGGTCCGGCA TTAACGCTAA ATCCTGATTT	1200
	TTTTATGTAT CAAGCATATG CGGCACAAGT AAATCGTGAA ATTGCATTTG TAGATGCAGG	1260
5	ATCAGATTTA ACGTTTGATT TGGAAACCAT TTAAACGAAA ATCGATGAAG TACAACCATC	1320
	ATTTTTTATT ATGAGTAATC CACATAACCC TTCAGGCAAG CAATTTGATA CGGCATTTTT	1380
	AACAGCTATT GCAGATAAGA TGAAAGCATT AAACGGATAC TTTGTCATTG ATGAAGCATA	1440
10	TTTAGATTAT GGTACGGCAT ATGACGTGGA ACTGGCACCA CACATCTTAA GAATGCGTAC	1500
	ATTATCAAAG GCGTTTGGAA TTGCCGGCTT AAGATTAGGT GTCTTAATTA GTACTGCTGG	1560
15	AACGATAAAG CATATTCAAA AAATAGAACA TCCATATCCA TTAAATGTAT TTACGCTAAA	1620
	TATTGCGACT TATATTTTTA GACATAGAGA AGAGACAAGA CAATTTTTAA CGATGCAACG	1680
	ACAGTTAGCT GAGCAGTTAA AACAAATATT TGATACACAT GTTGCAAGATA AAATGTCAGT	1740
20	GTTCCCATCA AATGCTAATT TTGTACTTAC TAAAGGCTCA GCAGCGCAAC AATTAGGACA	1800
	ATACGTATAT GAACAAGGAT TTAAACCTCG CTTTTATGAT GAGCCGGTGA TGAAGGGCTA	1860
	TGTAAGATAC TCAATTGCAA CAGCATCACA GTTAAAGCAA TTAGAAGAAA TTGTTAAAGA	1920
25	ATGGAGTGCA AAATATGATT TATCAAAAAC AACGAAACAC AGCTGAAACG CAACTAAATA	1980
	TTTCAATATC TGATGATCAG TCACCATCGC ATATTAATAC AGGTGTGGGC TTTTAAATC	2040
	ATATGTTAAC CTGTGTTACA TTTCATAGCG GTCTGTCATT AAACATTGAG GCACAAGGTG	2100
30	ATATTGACGT AGATGATCAC CACGTAAC TG AAGATATCGG CATTGTCATT GGCCAATTGT	2160
	TACTTGAAAT GATTAAAGAT AAAAAGCATT TCGTTCGTTA TGGAAACGATG TACATTCCAA	2220
	TGGATGAAAC ATTAGCACGT GTCGTTGTGG ATATAAGTGG GCGCCCATAC CTATCATTCA	2280
35	ATGCATCATT AAGTAAAGAA AAAGTTGGTA CGTTTGATAC GGAGTTAGTA GAAGAATTTT	2340
	TTAGAGCGGT CGTAATCAAT GCAAGATTAA CAACGCATAT TGATTTAATT CGTGGAGGTA	2400
40	ATACACACCA TGAAATTGAA GCTATATTCA AAGCGTTTTC CCGTGCATTA GGCATAGCGC	2460
	TAAC TGCAAC TGATGATCAG CGTGTGCCGT CATCGAAAGG TGTGATTGAA TGATTGTCAT	2520
	CGTTGATTAT GGATTAGGGA ATATTAGTAA TGTAAAACGC GCTATTGAAC ATTTAGGGTA	2580
45	TGAGGTGGTT GTCTCAAATA CCTCAAAAAT AATCGATCAA GCAGAAACAA TCATATTGCC	2640
	CGGTGTCGGC CATTTTAAAG ATGCGATGTC AGAGATAAAA CGATTAAATC TCAATGCAAT	2700
	ATTGGCTAAG AATACTGATA AGAAGATGAT TGGTATTTGT TTAGGCATGC AATTAATGTA	2760
50	TGAGCATAGT GATGAAGGCG ATGCATCTGG ATTAGGGTTT ATCCCAGGAA ATATTTTCGCG	2820
	TATCCAAACA GAATACCCAG TGCCACACTT AGGCTGGAAT AATTTAGTGA GTAAGCACCC	2880
55		

	AATTGCATAT GCGCAGTATG GGGCTGATAT TCCGGCAATT GTTCAATTTA ACAATTATAT	3000
	TGGTATTCAA TTCCATCCTG AAAAAAGCGG TACATATGGG TTACAAATTT TCGGTCAGGC	3060
5	AATACAAGGG GGATTTATAA ATGATTGAAT TATGGCCAGC GATTGATTTG ATTGGGTCAA	3120
	CAAGTGTGAG GTTAACAGAG GGTAAATATG ATAGTGAAGA AAAAATGTCA CGCTCGGCTG	3180
	AAGAAAGTAT TGCTTACTAT AGTCAATTTG AATGTGTGAA TCGTATTCAT ATCGTCGACT	3240
10	TGATAGGTGC TAAGGCACAG CATGCCCGAG AGTTTGATTA TATTAAGTCA TTGAGGAGAT	3300
	TAACAACCAA AGATATTGAA GTAGGTGGTG GCATTCGTAC GAAGTCACAA ATCATGGACT	3360
	ACTTTGCCGC AGGGATTAAT TATTGCATAG TTGGAACGAA AGGTATTCAA GATACTGATT	3420
15	GGCTTAAAGA GATGGCACAT ACATTTCCAG GTCGCATTTA TTTATCTGTT GATGCCTATG	3480
	GAGAAGATAT TAAAGTGAAC GGATGGGAAG AGGACACAGA GTTAAATTTA TTTAGTTTTG	3540
20	TGAGACGGTT ATCGGATATA CCTCTTGGCG GCATTATATA TACTGATATT GCTAAAGATG	3600
	GCAAAATGTC CGGACCTAAC TTTGAATTAA CTGGTCAATT AGTAAAGGCA ACGACGATTC	3660
	CTGTCATTGC TTCCGGTGGT ATTAGACATC AGCAAGATAT TCAACGATTA GCGTCGCTAA	3720
25	ATGTTACGC TGCTATTATA GGAAAGGCTG CACATCAAGC ATCTTTTGG GAGGGGCTAA	3780
	AATGATTAAA AAACGTATCA TTCCATGTTT AGATGTCAAA GATGGTCGTG TCGTTAAAGG	3840
	GATTCAATTT AAAGGATTAA GGGATATTGG GAATCCTGTT GATTTAGCAA TGTATTACAA	3900
30	TGAAGCGGGT GCTGATGAAT TAGTATTTTT AGACATCTCT AAGACGGAAG AGGGTCATAG	3960
	CTTAATGCTA GAAGTGATTG AACAGACAGC GTCACGCTTG TTTATCCCTC TTAGTGTAGG	4020
	GGGTGGGATT CAAAGTCTCG ATGATATTAC CCAATTGCTA AATCATGGTG CAGATAAAGT	4080
35	ATCATTAAT TCAAGTGCTT TAAAAATCC ACAGCTCATT AAACAAGCGA GTGATAAATT	4140
	CGGTAGACAA TGCATCTGCA TAGCAATTGA TAGCTATTAT GATCCTGAAA GAAAAGCACA	4200
40	TTATTGTTGT ACGACTGGTG GTAAAAAAT GACAAATATT AAAGTATATG ACTGGGTACA	4260
	GCAAGTAGAA CAGTTAGGTG CAGGTGAGCT CCTCGTTACA AGTATGGGAC ATGATGGTAT	4320
	GAAACAAGGC TTTGATATTG AACACCTAGC AAATATTAAG TCTCTTGTA ATATTCCAAT	4380
45	CATTGCTTCT GGTGGTGGTG GCAATGCACA ACACTTTGTA GAATTATTTG ATCAGACGGA	4440
	TGTTTCTGCA GGTTTAGCTG CAAGTATATT ACATGATCGA GAAACGACGG TTCAATCTAT	4500
	TAAAGAAGTG ATACGGCAAG GGGGTATAGC AGTAAGATGA CCAAATATAA AATTGATTTT	4560
50	AGCAAAGGTT TAGTGCCAGC AATTTTACAA GATAATCAAA CAAAACAAGT ATTGATGTTG	4620
	GGTTATATGA ACCAAGAAGC TTTTGATAAA ACGATAGAAG ATGGTGTGGT ATGTTTCTAT	4680

55

EP 0 786 519 A2

	AAAGATATTC ATGTAGATTG CGACAATGAC ACTATTTTAA TTGATGTCAT ACCAAATGGA	4800
	CCAACATGTC ATACAGGCAG TCAAAGTTGT TTCAACACAG AAGTTCCATT TTCAGTGCAA	4860
5	ACATTAGCGC AGACAGTTCA AGATAGTGCC CAATCCAATA ATGAAAAGTC ATATACAAAA	4920
	TATTTATTAA CAGAAGGTAT AGAAAAGATT ACAAAAAAAT ACGGTGAAGA AGCTTTTGAA	4980
	GTCGTAATTG AAGCAATTAA AGGTGACAAA AAAGCATTTG TAAGTGAAGT AGCAGATGAA	5040
10	CTTTATCATT TATTTGTCTT GATGCATGCG CTTGGCGTCG ATTTTTCAGA AATTGAGGCG	5100
	GAATTAGCGC GTAGACATCA TAAGCGCAAT AACTTTAAAG GTGAACGACA AAATATCGAA	5160
	CAGTGGTAAA GCAAGTATGG ACTAAGATAT AAGGAAAAGG ATCATGGCTT ATACACTTAC	5220
15	AAATATTGTG GAAAACGTGA CATTTTCAAG TTTAAAATAC GACACCAACA TATTTTAACT	5280
	ATGAATGCTG TGATGGTACT AAAGTTGCCA ACTCGTTATA GATAAGTAGT GGATAATCAC	5340
	AATACGAAAT CAAAAATAAT TATAAAAAGT AAATTGAGCA ACTCAGGAAT AGATGTCACT	5400
20	GTTAAAGATG TCGAAAAGTA TATGAATCGA TATAATGAAG TTATGAAGGG AAAAAATGGC	5460
	GAAAAAGCTA AAGAGTTATG TTTGTCGTTA CTACCTATTA ATATCATAGT TGTCTTTACA	5520
25	TTCTTTGTAT TTATACTATA AATACAAATA TATCTAGCCT GAAATAGAAA TGTCATAGCC	5580
	TATTTAAAAG ACAATCTCCA TTAGAACTAA GATATGCATC CCGAAAGTTA GACTAAAAAA	5640
	CTAACTTTAT GGGATGTATT TTTATGCTAA TCATCATAAA TTCGAGATTA AGTTAAAGGT	5700
30	AGTTCAGAG TAATTAAACA ATAAATWAAA AATAGTAGGA TACTTACTTT GAGGGAAGAA	5760
	AATTAAGTGT ATATATTTAG TTTAGGAACA AGTATTACGG TTTATCCTGA TACAATTATT	5820
	GTGGATGGGA TGATATTTTT AGGTTTAAAA TACGACACCA GCAAACATAA TAACTGTAAT	5880
35	AGCTCATAAA TCTCCCCATA TAGCTAATCT AAAAAAATAA TACATCATTG GAATTAAGCC	5940
	CCAAGCATGT AAATATTAAA AATCAAAAmA GATATmTGTA AAAaAGTTAC AATTtGCATA	6000
	ATTAAATTGT GTCTAATTAT TGAATAATTA AATTTTGCCA AATATAATAT TAATTAATAA	6060
40	TTTGaAATGA TTAGCGTATA CACTTTAAAT TCTCTTTGGA GAATATATTT TTAAATACA	6120
	AATGTAAACG CTTTCTCGTC AAATTAAACA ATAGAAAGGA TGGTCATTAT GAGTGCTTGG	6180
45	TTAAGTAAAT TATTTGAGTT TATTCCTCGA ATAATTATCA ATTTGTTTAT CTAAAATAAA	6240
	AAAATAGAGG TGCTGACAAT GATGAAAAGT CAAAATAAGT ATAGTATTCG TAAATTTAGT	6300
	GTAGGTGCAT CTCCATTTT AATAGCTACA TTAATTTTT TAAGTG	6346

50 (2) INFORMATION FOR SEQ ID NO: 238:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3775 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

	TGATAATGAT TCTCATTGTC ATACATCACG AAGGAGGCTA ATTAGTCAAT GAATAAAGTA	60
10	ATTAAAATGC TTGTTGTTAC GCTTGCTTTC CTA CTGTTT TAGCAGGATG TAGTGGGAAT	120
	TCAAATAAAC AATCATCTGA TAACAAAGAT AAGGAAACAA CTTCAATTAA ACATGCAATG	180
	GGtACAAC TG AAATTAAGG GAAACCAAAG CGTGTGTTA CGCTATATCA AGGTGCCACT	240
15	GACGTCGCTG TATCTTTAGG TGTTAAACCT GTAGGTGCTG TAGAATCATG GACACAAAAA	300
	CCGAAATTCG AATACATAAA AAATGATTTA AAAGATACTA AGATTGTAGG TCAAGAACCT	360
	GCACCTAACT TAGAGGAAAT CTCTAAATTA AAACCGGACT TAATTGTCGC GTCAAAAGTT	420
20	AGAAATGAAA AAGTTTACGA TCAATTATCT AAAATCGCAC CAACAGTTTC TACTGATACA	480
	GTTTTCAAAT TCAAAGATAC AACTAAGTTA ATGGGGAAAG CTTTAGGGAA AGAAAAAGAA	540
	GCTGAAGATT TACTTAAAAA GTACGATGAT AAAGTAGCTG CATTCCAAAA AGATGCAAAA	600
25	GCAAAGTATA AAGATGCATG GCCATTGAAA GCTTCAGTTG TTAACCTCCG TGCTGATCAT	660
	ACAAGAATTT ATGCTGGTGG ATATGCTGGT GAAATCTTAA ATGATTTAGG ATTCAAACGT	720
	AATAAAGACT TACAAAAACA AGTTGATAAT GGTAAAGATA TTATCCAACT TACATCTAAA	780
30	GAAAGCATT CATTAAATGAA CGCTGATCAT ATTTTGTAG TAAAATCAGA TCCAAATGCG	840
	AAAGATGCTG CATTAGTTAA AAAGACTGAA AGCGAATGGA CTTCAAGTAA AGAGTGGAAG	900
35	AATTTAGACG CaGTTAAAAA CAACCAAGTA TCTGATGATT TAGATGAAAT CACTTGGAAC	960
	TTAGCTGGCG GATATAAATC TTCATTAAAA CTTATTGACG ATTTATATGA AAAGTTAAAT	1020
	ATTGAAAAAC AATCAAAATA ATTAAGGAGT TTTACGATGC TACTTAAACC AAAATACCAA	1080
40	ATCGTTATTG CTGGTTTATG TCTTGCAATA GTAGCTATCT TAAGTTTAAT GATTGGAAAT	1140
	ACGCTTGTGT CACCAGGTAC GGTGATACAG GCGTTATTCA ACTTTGATAG TGAAAACGAT	1200
	TTACATGATG TTGTCAC TGG TGCACGGGCG TCGAGAACAA TCATTGCGTT ATTGACTGGT	1260
45	GCTGCCCTTG CTGTCTCAGG TTTGTTGATG CAAGCACTTA CACGAAACCC AATAGCCTCA	1320
	CCAGGGCTTT TCGGTGTCAA TGCAGGCGCA GTATTTTTTG TCATTTTTAG TATTACATTT	1380
	ATCCAAATTC AATCTTTTAA AATGATTGTA GTTATTGCAT TTTTGGGGGC TATTGTTGTT	1440
50	ACTGTATTAG TTGTTGCACT AGGTATGTTT AGACAAACAC TATTCTCACC TCACCGTGTC	1500
	ATTTTGGCAG GTGCTGCGAT TGCATGCTA TTTACAGCCT TTA CTCAAGG CATACTTATT	1560

55

	AATATTTGGG ATATCCCATG GATTATTCCG CTTGTATTGA TACTTATTTT AATTGCATTT	1680
	AGCATGGCTG CACACATCAA CATCTTGATG ACAAGTGACG ACATTGCAAC CGGCCTCGGT	1740
5	CAAAACATAA AATTAATCAA ATGGATGATT ATTATGCTCA TCAGTATGTT AGCCGGTATT	1800
	TCGGTAGCCG TAGCTGGATC AATCGTCTTT GTGGGTCTTA TCGTACCGAA TATTAGCAAA	1860
	CGATTATTAC CACCAAATA TAAGTATTTA ATTCTTTTCT CTGCATTAGC TGGAGCAATC	1920
10	CTAATGATCA TTTCAGACAT TGTGCTCGT ATAATAATTA AGCCACTAGA GTTGCCTATC	1980
	GGTGTGCTTA CCGCTGTCAT TGGCGCTATT GTCTTAATCT ATATTATGAA GAAAGGACGT	2040
	CAACGCTTAT GACCGAAAAG ATTAATAAAA AAGACAATTA CCATCTCATC TTCGCGTTAA	2100
15	TCTTTTTAGC CATCGTTTCA GTGGTAAGTA TGATGATTGG TTCAAGCTTT ATACCATTAC	2160
	AACGCGTACT GATGTACTTT ATAAATCCAA ATGACAGTAT GGATCAATTC ACTTTAGAAG	2220
	TATTACGCTT ACCTCGCATT ACCTTGCGA TTTTAGCAGG TGCCGCACTA GGAATGAGTG	2280
20	GTTTAATGTT GCAAATGTA TTAAAAATC CAATTGCCTC ACCTGATATT ATCGGTATCA	2340
	CAGGTGGTGC TAGCTTAAGT GCTGTTGTCT TTATTGCATT TTTCAGCCAT TTAACAATAC	2400
25	ATTTACTTCC ACTATTTGCA GTATTAGGTG GCGCAGTTGC AATGATGATA CTATTAGTGT	2460
	TTCAAACGAA AGGACAAATA CGCCCGACAA CACTCATAAT CATCGGTATT TCGATGCAAA	2520
	CGTTGTTTAT TGCCTTGTG CAAGGATTAC TCATTACAAC GAAGCAATTA TCTGCTGCCA	2580
30	AAGCTTATAC ATGGCTAGTC GGAAGTCTTT ACGGTGCTAC GTTTAAAGAT ACAATCATTT	2640
	TGGGTATGGT TATTTTAGCT GTTGTGCCGT TGTATTCTCT TGTTATACCA AAAATGAAAA	2700
	TATCTATACT TGATGACCTT GTAGCGATTG GCTTAGGCTT ACATGTACAA CGTATGAAAC	2760
35	TAATCCAATT AATCACTTCT ACTATACTCG TATCTATGGC AATCAGTTTA GTAGGTAACA	2820
	TTGGGTTTGT CGGTTTAATC GCACCACATA TCGCGAAAAC AATCGTTCGC GGAAGTTATG	2880
	CTAAAAAGTT ACTAATGTCA GCAATGATTG GTGCCATATC AATTGTTATT GCAGACTTAA	2940
40	TTGGGCGTAC CTTATTCTTG CCTAAAGAAG TGCCAGCAGG TGTATTTATT GCTGCTTTTG	3000
	GTGCCCCATT CTTATATAC TTATTATTAA CCGTGAAAAA GTTATAACGA TATTATTAAA	3060
	ACAAAATGAC CTCACAACGA AGTTAGCTAA ATGATTCACT TAACCTAACG TTGCGAGGTT	3120
45	TTTTTATACA TATAGTTGTT GTTATTGTTA ACAAGCGTCG ACTTTCTTAA TTACATATTA	3180
	ATACTTTATA TACAAATAAC ACCGACTCAT ATTCTATAAT ATCAATCAAT ATTCTTCGAT	3240
50	TTTTCAAATA TCGATAACTA TTTCTTATTT AAATATAGTG TTTGATAATG TCATTTATTC	3300
	AAAAACACAA ATTTTAATAA AAATATCATA TTATTTTTAA TTGTAAATTA TGGATTATTT	3360

ATATAATAAC CATGACAAAA TTAGAGATTT TATAATCATT GAAGCATATA TGcTTCGTTT 3480  
 TAAGAAAAAA GTCAAGCCTG AAGTCGATAT GACTATAAAA GAATTTATAT TACTGACTTA 3540  
 5 TTTATTTTCAT CAGCAAGAAA ACACACTTCC ATTTAAGAAG ATTGTTTCAG ATTTATGTTA 3600  
 TAAACAATCG GATTTAGTAC AGCATATAAA AGTACTGTG AACATTCAT ATATTAGTAA 3660  
 AGTTCGAAGT AAAATTGATG AGCGTAATAC TTACATTTCA ATATCTGAAG AACAAChAGA 3720  
 10 nAAAATTGCA GAChTGTTA CATTGTTTGA TCAATCATT AAACAATTTA ACCTT 3775

## (2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 1361 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

AGAAAAATTA GCCTACCTAT GCAAGTTGTT ATTGCATTAG TGTTAGGTGT TGTCGTAGGA 60  
 25 CTTTATTAT ATGGACAAGA AAACGTAGCA AATTACATTA AACCATTGG TGACGTATTT 120  
 TTAAATTTAA TTAAATGAT CGTTATACCA GTTGTATTTT GCTCACTAGC GCTTCTATT 180  
 TCGAACGTTG GGAATCGAA AACTGTAGGG CGTTATGGCT GGaAAACAAT TTTATACTTT 240  
 30 GAAATTATTA CAACAATCGC AATAGGTTTA GGGATTATCT TCGGTAACTT ATTTAAACCA 300  
 GGTGCTGGAT TAGACCCAAC AAAATTACCT AAAGGTGATA TTTCTAAATA TCAATCAACT 360  
 GCACATGCAG CAGAACAATC TACATATGGA AATCATTTTA TTGATACCAT TGTACATATT 420  
 35 ATTCCGACAA ACTTTTTTGA AGCTTTAAAT AAGGGTGAAT TATTACCTAT TATCTTCTTC 480  
 GCAGTATTCT TTGGATTAGG ATTAGCTGCT GTAGGTAAAA AAGCAGAACC AGTTAAAGAA 540  
 TTTTAAGCG GATCGCTTGA AGCTGTGTTT TGGATGATTA ATAAAATTTT AAAATTAGCA 600  
 40 CCACTTGGAG TGTTTGCATT CATTTGTACT ACAATTATTA CATTTGGTGC ATCCGCATTA 660  
 TTACCACTAT TAAAATTAGT ATTAGTTGTT GTCTTTGCAA TGGTGTTCTT TGTATTCGCT 720  
 ATACTAGGAC TAGTTGCATG GATGTGTGGT ATTAATATCA TGAATATTAT TAGAATCTTG 780  
 45 AAAAGTGAAT TGCTTTTAGC ATTTTCTACA TCAAGTTCGG AAGCTGTACT TCCTGTAATG 840  
 ATGAAGAAAA TGGAAACTT CGGTTCTCCA AAAGAAATTA CTTCTTTTGT TATACCAATT 900  
 50 GGTATACGT TTAACCTAGA TGGATCAGCA CTTTATCAAT CTATTGCAGC ATTATTCGTT 960  
 GCACAGATGT ATGGAATGCA CTTAACATTA TCAGAGCAAA TTGTGTTGAT GTTAACATTA 1020

55

ACATTAGGTG CCATGGGCTT ACCGGCACAA GGTTTAGCAT TAATTATTGG TGTGACCGT 1140  
 ATCTTAGATA TGGTACGTAC ATGTGTAAAC GTTATTGGTA ATGCATTATC AACAATCGTT 1200  
 5 ATAGCTAAAT GGGAAAACGT ATATGACAAA GCAAAAGGTC AAGAATATTT AAAATCAATT 1260  
 TAAAAAATAC TATCTGACAT TTAArGnCCC TTACAACCTT TGGTTgTnAG GGCTnTTTTA 1320  
 TGTCATGCGT CTTAAAGCCA GGCCGTATAn CGGTAAGCGT A 1361

10 (2) INFORMATION FOR SEQ ID NO: 240:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1489 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

TCAATATGTA AATCGGCGAT TGTCCCTATC tTCATACCCT TACCTTCTTT ACTCAAATTA 60  
 AAATGTTATT TATTATGCCA TaGCTTATCT AATATATATA GTTAACTGCT TCATTTAGAT 120  
 25 GATTATTTTA TATTTTTGCA TAAAACTTA TATCTTTTCA AAAAAATCGA TAAGTTTtag 180  
 TTATCATACC CTTACCTATC AAATGTTTTT TCTTATATTT AAAAAAATAA TTGCTTTATT 240  
 AAATGGATTT CTtTAGTATT TATAATTAAG AAAACGCTTA CACACAACCT TTTTATTtGC 300  
 30 TTTATCCTGA GGAGGAAAAT TATGGCAAGA AAATtGCATA GAGAGTTGAA TAACAGACAC 360  
 ATCCAATTAA TAGCAATTGG GGGCGCAATT GGAActGGT TATTCCTAGG ATCAGGTCAA 420  
 ACAATATCTT TAACTGGTCC ATCACTGTTA TTCACATACA TGATTATTGG GGTtGTACTA 480  
 35 TTCGCTTTTA TGCGCGCATT AGGCGAATTG TTGTTGAGCA ATACAAGATT TAATTCATTT 540  
 GTTGATATTG CAAATGAATA TTTAGGCCCT TTTGGTGGCT TTGTCATTGG CTGGACTTAC 600  
 TGGTTATGTT GGATTGTATC AAGTATGTCA GACCTAACTG CGATGGGACA ATACTTTGCA 660  
 40 TTTTGGTATC CACAAGTCCC AAATTGGATT ACCGTGCTAT TTATTGTTTT AATCTTGATT 720  
 AGCTTCAACT TATTAGGTGC CAGATTATTT GGTGAActGG AGTTTTGGTT CTCGATTATT 780  
 AAAGTTGTCA CAATTATTGC GATGGTTATC GTTGGTCTTG TATTAACTTT TTTCTCATT 840  
 45 AAAACACATT ATGGACATGC ATCATTCAAA AACTTAATCA GTCACGGTGG CATGTTCcCT 900  
 GGTGGAaCAT TTGGTTTCTT AATGTCAATC CAAATTGCTG TATATTCAAT CATTGGTATT 960  
 50 GAACTTATAG GTGTAACTGC TGGTGAAaCG AAAGATCCTG AAAAAACCTT ACCGAAAGCA 1020  
 ATTAATAATG TACCTATCCG TATTTTATTA TTCTATATCG GTGGTCTATT AGTAATTATG 1080

55

TTAATCGGCG TACCATTTC AGCAGGTGTC GTTAACTTTG TCGTGCTAAC TGCCGCGGCC 1200  
TCTGCTACAA ATAGTGGTAT CTATTCGAAT AGTCGTATCT TATTCGGA CTGACAACAA 1260  
5 GGGTTAGGTC CTAAAGTTTT AAATAAAACG AATAGTCATG GCGTGCCTTA TTTATCAATG 1320  
TTAGTTTCAT CAATTGCATT ACTTATAGCA GCCTTGTTAA ACTACATTTT CCCTAATGCA 1380  
ATTCAACTAT TCAATACGT TACAACGTTA tCAACTGTGT TGTTTTTAGT TGTtTGGGCA 1440  
10 ATGATnATTG TCGCTTATCn AATGTATTTG GAAAAAGCAT CCTGAGGCA 1489

(2) INFORMATION FOR SEQ ID NO: 241:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 5000 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

TTTTCCATCA TrATcACCAT TTAAGACAAT AGCTGTATGA CCATATCCTC CACCAAATCT 60  
25 TCCACTAAAA ACCACTAAGT CCCCAGGTTT AGCTTTGAAA GTTGGTGTGT TGTGATAAAT 120  
TTAGCTTCA CTATTAAAT TATTTGCATA TGGTATATCT TTAGCTCCAT ATCCTTTTAA 180  
TCCAKGACCA TAAAGAKGAT TCCAATATAC ATTAATAAA TCGAAACATT GCCATCCATA 240  
30 ACTACCGTCG AAATCCCATC CTCTATTCTC TAATGTTTTT AAATAATTTA ATGTACTGCT 300  
ATTACTATTA CTTTTATTAT TTGAAGACAC TGTTTTTGGT TTTGGTTCTA CTAATGGTGT 360  
CATTGGCACT TTTAATTTTT GACCAATAAA TATTAAATTA GGATTTGCTA TATTATTTGT 420  
35 ATTTTGAATA TTTGAAACTG TAGTTTTGTA TTTAATGCT ATAGCACTAA GTGTGTCTCC 480  
TTTTTTTACA GTATAGATT GTGTTTTTGG AGCTTCTCTA AACTGTAGT AACCAGGTA 540  
ATTATTAGTA ACTTTATTCG TTTCTTCTG ATTAGAATT TGAGCTTCCA AGTTTGCAAT 600  
40 TTTAATTTCT TTAGTAAGTT CATTGTTATT AATAACTAGA TTGTTACCTT GGCTTGAGTT 660  
TTTCGGAGTA TTTGAAATCT TTATATCTTG ATTAATTTCA TTTCCGTTTG AAATGCTGA 720  
TTGTTGTCT AACTTTAAAC TTGTGTCGGA TGTTTTAAAC GCACCTTCAT TTTTTATTTT 780  
45 GTCTTTTGTC GTATTTTTAT TAGCATTTAA CTCTGATTC GCGAATACAT TTTGCTCATA 840  
CCCTCTTGTA AAATCTTTAG ATTTATCAAT TTCATCTGCA TATGCTTTGT TCGACATACC 900  
50 CAATGCCAAA AACATACCTA TTGAAATTGA CAAAATTCCA ATACTAACTT TTCTAATTGA 960  
ATAGCGTACT TTACTTTGTT GTTTATTCAT GATGAAACAC TCCTTACAAT AATATACAAA 1020

55



GTAGGAGGTA AAATAATTAA CTTGTCTTTC CAAAATATGA AAAGTGTACT AAAaTTCATC 1140  
 GCACGACAAA TAGCCCATT TCCGATACTTT TATAAAGTAT GGAaTGGGCT ATAGCCATT 1200  
 5 ATATCATCTT TTAACCTTAT TTATTAACAG TTAATAATGA TTCATAAATA CCTGCCTCTT 1260  
 TAGCAGCTTC AATTAATGTT GAACCAATTT CTGAAGGTGT TGCCGCTGTT TTCACACCAC 1320  
 AACTATTTAA TGTTTTAATT TTCTCTTCAG CAGTACCTTT ACCACCTGAA ATGATTGCAC 1380  
 10 CAGCATGTCC CATACGTTTT CCAGGAGGTG CTGTTTGTCC ACCGATAAAG CCTACAACCTG 1440  
 GTTTTGTCTT ATTCGCTTTA ATCCATTTCAG CTGCTTCTTC TTCAGCCGTA CCACCGATT 1500  
 CACCAATCAT AACAACTGCT TTCGTTTCGT CATCTTCATT GAATGCTTTT AAAACATCAA 1560  
 15 TAAAGTTTGT TCCGTTGACT GGGTCTCCAC CAATACCAAC AGCTGTAGTT TGACCAATAC 1620  
 CTTCTTCAGT CAATTGGTGC ACTGCTTCAT ATGTTAATGT ACCTGAACGA GATACTACAC 1680  
 CAACATGACC TTTTTTGTGA ATATAGCCAG GCATAATACC AATTTTACAT TCATCTGCTG 1740  
 20 TAATCACACC TGGACAGTTC GGACCAACTA AACGTGTTTT TCTACCTTGT AAGTAGCGTT 1800  
 TAACCTTAAC CATGTCTAAT ACAGGAATAT GTTCAGTGAT ACAAATAACC ATATCTAAGT 1860  
 25 CTGCATCAGC TGCTTCTAAA ATTGAGTCTG CAGCAATGG TGCTGGAACG TAAATGACTG 1920  
 AAACCGTTGC CCCAGTTTCA TTTTGTAGCTT CTTCAACAGT GTTGAAAACA GGAACGCCTT 1980  
 CAACAACTTG ACCACCTTTA CCAGGCGTCA CACCTGCTAC TATTTTCGTA CCATAATCAA 2040  
 30 GCATTTGTTT TGTATGGAAA AGGGCAGTAG ACCCTGTAAT ACCTTGTACC ATTACTTTAG 2100  
 TATTCTTATC TATAAATACA CTCATCTTAG TGCTCCCATC CTTTCCTTAT GCTTCTTTGA 2160  
 CTAGTTTAAC AATTTTTTGT GCACCTTCAG CCATTGTTGC TGCTGGTTCA ATTGCTAATC 2220  
 35 CTGAGTCTTT TAAGATTTTT TTACCTAACT CAACATTGTT ACCTTCTAGG CGTACAACCTA 2280  
 GTGGTAAAGT TAAATCTACT TCTTTTACAG CTTCAACGAT ACCTTCTGCG ATAACATCAC 2340  
 ATTTTATAAT GCCACCGAAA ATGTTTACAA AAATACCTTT AACATTTTCA TCACCTAAAA 2400  
 40 TGATTTTAAA TGCTTCAGTT ACTTTTTTCTC TAGTAGCGCT TCCGCCTGCA TCTAAGAAAT 2460  
 TGGCTGGGTT TCCACCGAAA TGATTAATCG TATCCATTGT TGCCATGGCT AAACCTGCAC 2520  
 CATTAAACCAT ACATCCGATG TCACCATCTA ATGCAATGTA TGATAAATCA TGTTTAGACG 2580  
 45 CTTCAATCTC TTTCCGATCT TCTTCTTCTA AATCACGTAA TTCTACAACA TCTTTATGTC 2640  
 TGAATAATGC ATTATCATCA AAATTAATTT TAGCATCTAA TGCCAATACA TCACCATCAG 2700  
 50 CTGTTGTAAC TAATGGGTTG ATTTCTACGA TTGAACAATC TTTTCAATG AATACATTAT 2760  
 AAAGTGCTAA TAAGAATTTA GCAGCTTTGT TAACAGATTC TTTAGGAATA TTAATATTAA 2820

55

	AGATCTTTTC AGGAGTCTTC GCAGCAACTT CTTCAATCTC AGTGCCCCCT TCTTCAGACG	2940
	CCATCAATGT TACTTGGTCA GTCGCACGAT CAATAACGAA TCCAACGTAA TATTCTTTTT	3000
5	GAATAGCACA ACCTTCTTCG ATATATAAAC GCTTAATTTT TTTACCTTCT GGACCAGTTT	3060
	GATGTGTAC CAAAGTTTTT CCTAATAATT CTTTTGCATA TGTTTCTACC TCAGATAAAG	3120
	ATTTAGCAAT TTTTACTCCG CCTGCTTTAC CTCTACCTCC AGCATGAATT TGTGCTTTTA	3180
10	CAACATAAAC ATCAGAATTT AATTCTTTTG CTTTCTCCAC CGCTTCTTCA GCAGTAAATG	3240
	CTACTCGTCC TTCTGGAACT GCAACGCCCA TTGAACGAAA TATTTCTTTA CCTTGATACT	3300
	CGTGATATT CATCTTCCAT CCTCCTGTTA CTTAGGTTAA GTTCCCTTAC AATTATAAAA	3360
15	AATGTAAGCG CTATTGTAAA CTTAAATGCT ACTTTTTTAT CATTTAATTG AATTTTACGA	3420
	TTTACAGTAA CGATTTTATA GGTTCAAAGC TTTTCTATG CTCTTTCATA ATGCCAATAT	3480
20	CATCGATTGC TAGTAAATGT TGTTTGGTAC CGTAACCCGC GTTTTTTTTCA AAACCATATT	3540
	CAGGATAATC TTTAGATAAC TGTGTCATAT AATCATCAGC AAAACCTTT GCCATGATAC	3600
	TTGCAGCTGC AATGGACACA CTTCTTGCAT CACCCTTGAT TAAAGATACT TGAGGCAGTG	3660
25	CATTATCAAG CGTCATCGCG TCTATCAATA AATGCGTTGG TTGTACTGAT AATCCATCAA	3720
	TAGCTCGCTG CATGGCGATT TGAGTAGCTT TATAAATATT AAATTCATCT ATTCTTnCA	3780
	GLGTCGCGAT CCCATATGCA AAAGCAGTAA CTTTATTTTT TAGTGCTTCA TTTAATTCTA	3840
30	GACGTTTTCGT AACAGGTACT TTTTCGAGT CATCAAGGCC CAAATAATTG TGATTTGAAT	3900
	TTAAATTTGT TGCGCATGCA ACGACTGGAC CTGCTAAAGG TCCTCTTCCA ACTTCATCAA	3960
	TCCCACAAAT AATAGCATTG GGATGCTCTT TTAATATTTT ATTTTCAAAG TAAGTCATTT	4020
35	CAACATACTT TTCTTTTAAA GCTTGTTCTT TTTCTAACGC TTTTCTGCGC CTAGCTATGG	4080
	CATTTTGAAC ACCTTTTCGC TCATCTAAAA AGCATTTCATG ATTTTCTAAT TCTTCTATTG	4140
	TATTAACCGC ATTAATCAAC TGCCTAACTT CTTTAATTGT TAGCGTCATT TGCTAATTCC	4200
40	TCAGTCATAT CTTTAAAAAT ATCAAAACAA TAATTTCTTA TTTTAGCATT TCGAATATCA	4260
	TAAATAATCA GTTCAATGAC TGCTTCGTAA TCAATTTTCA TACCACGTCG AATTAAGCCC	4320
	ACGTTTTTTC CCTATCGCAT CAAACCACGC TATGATTTCT GCATCTTCAG GAACTTCAAT	4380
45	ATTATAATGT GACTTTAATC GCGCTAAATC ATTTTGAATT AAAAAGTTTA ATCCATAGAT	4440
	GGCAACTTCA TCTAAGTGCA CAATACTATC TTTTATCGCA CCAGTTAAAC TCAACTTCTT	4500
50	ACCGACTTCT TCATCTTCAA ATTTAGGCCA AAGTATCCCT GGTGTGTCTA ATAGTTGTAA	4560
	TGCATTACCA ACTTTAATCC ATTGTTGTTG TTTGGTCACA CCTGGTTTAT TACCAGTCTG	4620

55

AACGATCATT GCTCTTATCG CTCTAGGTTT AAGTCCTTTC GCTTTTTTCGC GTTCAAATTT 4740  
 TTCAGCAGTC GCCTTAATTG CTGCAGCTTC CACTTTCTTT AAATTTTAC CGTGCTTAGC 4800  
 5 ATCCACTGAT ACAGGATAgT AtCCTTTATC AATAAAAAAT tGTTCCCATTT TTGACATCTC 4860  
 ATTTAAATTA GACATATCTT TTTTATTTAA TATAACAACA CGTGGTTTTT GGTAAATAAC 4920  
 TTCATCTATC ATAGGGTTTC TTGAATAATA TGGAATTCCT GCATCTACTA GTTCAAACAC 4980  
 10 TACATCTACT TTTTTTAATT 5000

(2) INFORMATION FOR SEQ ID NO: 242:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1700 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

AACCCGnAGA CGAAACTnCA TAGTTGCCTG TTATGTTGAT TAGTGCGTTT GTAGCTATTT 60  
 25 nAAATCAAAC ATTATTAAAT ACAGCGTTaC CTAGTATAAT GAGAGAATTA AATATCAATG 120  
 AAAGTACATC GCAATGGCTA GTTACTGGGT TTATGCTTGT TAATGGCGTC ATGATACCTC 180  
 TGACGGCATA TCTAATGGAT AGAATTAAAA CTAGACCTTT ATACTTAGCG GCGATGGGGA 240  
 30 CATTTTTATT AGGTTCTATT GTTGCGCCT TAGCTCCGAA TTTTGGAGTT TTAATGTTAG 300  
 CTCGTGTAAT TCAAGCGATG GGTGCAGGCG TACTTATGCC CTTAATGCAA TTTACGTTAT 360  
 TTACATTGTT CAGTAAAGAA CATCGAGGTY tGCAATGGG ACTAGCAGGT TTAGTAATTC 420  
 35 AATTTCACC AGCAATAGGA CCTACAGTTA CAGGATTAAAT TATTGATCAA GCGAGTTGGC 480  
 GAGTfCCATT TATTATAATT GTAGGAATTG CTATACTTGC CTTTGTtTTC GGTTTGGTTT 540  
 CAATCTCGAG TTACAATGAA GTGAAATATA CGAAATTAGA TAAGCGTTCA GTAATGTATT 600  
 40 CAACTATTGG GTTCGGGTTA ATGCTATACG CATTTAGTAG TGCAGGAGAT TTAGGATTTA 660  
 CAAGTCCAAT AGTAATAGGT GCGTTGATAT TAAGTATGGT TATTATCTAT TTATTTATAC 720  
 GTAGACAATT TAATATTACT AATGCACTTT TAAATTTAAG GGTTTTTTAA AATAGAACAT 780  
 45 TTGCATTATG TACGATTAGT TCAATGATTA TAATGATGTC AATGGTTGGA CCTGCGCTGC 840  
 TTATACCGCT ATATGTTCAA AACAGTTTAT CTTTATCTGC CTTGTTATCA GGAATTGTTA 900  
 50 TCATGCCTGG TGCAATAATA AATGGTATTA TGTCAGTTTT TACAGGTAAA TTTTATGATA 960  
 AGTATGGTCC TAGACCATTG ATTTATACTG GTTTTACAAT TTTAACAATT ACTACAATTA 1020

55

GAATGTTTTTC AGTTTCTTTA CTCATGATGC CGATAAATAC TACAGGAATT AATTCTTTGA 1140  
 GAAATGAAGA AATCTCACAT GGCACGGCTA TTATGAACTT TGGTCGTGTA ATGGCTGGTT 1200  
 5 CACTAGGCAC AGCTTTAATG GTTACATTAA TGAGTTTGG TGCAAAAATA TTTTATCTA 1260  
 CATCGCCATC GCATTTAACT GCAACTGAAA TTAAACAGCA ATCCATTGCT ATAGGGGTGG 1320  
 ATATCTCATT TGCTTTTGTA GCTGTGCTTG TTATGGCAGC TTATGTGATA GCACTTTTTA 1380  
 10 TAAGAGAACC TAAAGAAATA GAAAGTAATA GAAGGAAATT TTAAAATAAT TATAGTAGTT 1440  
 GGTCTATTTA AAATAATAGG CTAAGTCTT TTTTATTTA ATAAAAAGTT TTATACTTTT 1500  
 AGTGATAGAC TAAGCAAAAA TTGTTATTTG CTATGATGTA GATGTCTTAA AATGATTAAG 1560  
 15 GGGGATTTGC TTTGTTAACG GTAGATCAAG TGAAAGAATT GGTAGGAGAA ATTAAAGATC 1620  
 CTATTATAGA TGTGCCTTTA AAAGAAACAG AAGGTATTGT TGAnGTTTCT ATTAAGGGAG 1680  
 20 AAnAAGAACA TGTGAGTGTT 1700

## (2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10146 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

TGCATCAACT TTCAAACAAT AAATCATCAC AATAACCACA CCTAATTCAA CACTTTTCAA 60  
 ACATAAGTAT TGACACATTG AGCAAAATGA TTTTAAATG TAACTAATAC AGTTACAATT 120  
 35 ATGAGGTGAG AAACATTGAA TTTAGAATTT AACATTGCCG TGCATGTATT AGCTTTTTTA 180  
 ACTAAGCATC aTTCAGAAAA ATTCAATAGT AGTTCATTAG CAGAAaTTAaC TTGTTTAAAT 240  
 CCTGTTCAAT TACGACGCGT GACGACTCAA CTTGTGCGATT TAAAAATGAT TGACACAATA 300  
 40 CGAGGTAAAG ATGGCGGTTA TTTAGCAAAT GATCAAAGTG CTGATGTCTC TCTAGCAACA 360  
 TTATaTAAAC ATTTTGTCTT AGAGAAAGAA CACCACACAC GTCTATTTAC TGGCGACGAA 420  
 GGCAGTCACT GTCAAATTGC TCGTAATATT GCAACTACCA TGTCACATTA TCAGCAAGAC 480  
 45 GAACAGAATA TCATTATTAA TTTTATAAT GGAAAAACAA TCAAAGATGT CATTGAAGAC 540  
 ATTCAAAAGG AGGATTTATG TCATGAAAAC ATATGATTTA ATTGTAATAG GATTTGGGAA 600  
 50 AGCTGGTAAA ACTTTAGCGA AATACGCTGC ATCAACAGGT CAACACGTCG CAGTTATCGA 660  
 ACAATCTCCG AAAATGTATG GAGGCACTTG TATAAACATA GGATGTATAC CTTCGAAGAC 720

55

	TGTTGTCAAT GCGCTAAACA ATAAAAATTA CCACTTATTA GCAGACGATA ACAACATTGA	840
	TGTACTGGAT TTAAAGCGC AGTTTAAATC TAATACTGAA GTGAATTTAT TAGATCAACA	900
5	TGACGATATC GTCGATAGTA TTA CTG CACC TCATATCATT ATTAATACCG GTGCTACCTC	960
	TGTCATTCTT AACATTAAAG GCCTTGATCA AGCTAAACAC GTCTTCGATT CGACAGGTTT	1020
	ATTAAACATT AGCTATCAAC CTAAGCACCT CGTTATTGTA GGTGGCGGTT ATATCGCCTT	1080
10	AGAATTTGCT TCAATGTTTG CGAATTTAGG TAGTAAGGTA ACAGTATTAG AACGTGGCGA	1140
	ATCATTTATG CCACGCGAAG ATCAAGATGT CGTTGCATAT GGTATTACTG ACTTAGAAAA	1200
	TAAAGGCATT GCATTGCATA CAAATGTTGA AACGACTGAA TTGTCATCTG ACAATCATCA	1260
15	TACAACAGTC CATACCAACG TTGGTAACTT TGAGGCTGAT GCAGTACTTT TGGCTATCGG	1320
	GCGCAAACCG AATACGGATT TAGCTTTAGA AAATACTGAT ATCGAATTAG GCGACAGAGG	1380
20	CGAAATTAAA GTCAATGCTC ATCTTCAAAC AACTGTGCCG CATATTTATG CTGCAGGTGA	1440
	TGTTAAAGGC GGA CTTC AAT TTACGTATAT ATCTTTAGAT GATTATCGAA TTATCAAATC	1500
	AGCGTTATAT GGTAATCAGT CACGTACGAC TGACAATAGA GGCAGCGTGC CTTATACAGT	1560
25	ATTTATAGAT CCACCATTAT CACGTGTTGG ATTAAGTAGT AAAGAAGCTG CCGCTCAACA	1620
	TTATGATTAC ACTGAACATC AACTTTTAGT AAGTGCTATA CCTCGTCATA AAATTAACAA	1680
	TGATCCAAGA GGT TTATT TA AAGTAGTCAT TAATAATGAA AATAATATGA TTTTAGGTGC	1740
30	TACATTATAT GGTAAGCAAT CTGAAGAATT AATTAATATA ATTAACTTG CGATTGATCA	1800
	AAACATTCCA TATACCGTAT TACGAGATAA TATTTATACG CATCCTACGA TGGCCGAATC	1860
	ATTTAATGAT TTATTTAATT TCTAGACAAA ACATAAAAAAC CTGGTGGCAC GCATTGAATG	1920
35	ATGCTGCCAT CAGGCTTTAT TGTTGTGCTT TTCGCTTTTC TAATTTTCT TTAAGCTTTC	1980
	TATCTGTGTT TTCTTTACGA CGTTTACGTT CTTTCATGTCG TTTTCTTAAA CGCTCTTCTT	2040
	CTTCAGGATC ACGTGGTTTC TTTAATTGTT GAGAACTTT TTCGATTAAAT TCTTCTTCAG	2100
40	TAAGCGCagc CAGTGGGCGG TTATTAACAA AAGTGAATGT TTTTCGGCGT CCAGGTCCAC	2160
	AATAAGATTG ACAACCTATC ACGATTTCAG CATCGGGATC TAATTTTCC AACTTCTTTT	2220
45	GTAACGTTCT TATATTGACT GCCTGACATT CATCACAAAT AAGGAATGTA TTTTTCATAT	2280
	TGCTACCCAC CTTTCTTTAT CATATCTATA TCGTCGATTT CATTAATTTT TTCGTTAACT	2340
	CTATCTATTT TACTCTTTTT AATATTTTTT TCAAGATACG TAACACGGCT GaCAATAAAA	2400
50	AATGGAGCAT TTATCTTCTA ATTAAATTAG ATGaTTGCTC CCCTATCAAA TCATTTATTG	2460
	CCCATGATAA ATATTAAATT TTAATGGTTT AATACCATGT TTTGTCCATT CATCATAAAT	2520

55

	TGATGTTTTA GATGCGCCGT GATACTTTTC AGCAATATCA CACAAATATT TTAGCTTTTC	2640
	AGTTTCTATA TCAACTGTAG CTTCTTTATC CATACGTTGA ATAATTGTAC GATTCTGACG	2700
5	CACCATCTTT TGCACACCTT TAATGTTATT TGTTTTAAAA GCATGAATAA GTTTTTCAAC	2760
	ACAAAGATGT GAATCTTCTA AGAAGTCACC GTAAAATGAA GGATCTGATT TCAAACGTTT	2820
	CACTTCGCTA ACAAAGTGTG GTGATGACGC CGGTGAGCCA GTCCAACCGA TAAGTACTTC	2880
10	CATATTTTCA GGTGCTTGTA ATGGTTCGAT GTGCAATCCA GGCCAGTTTT TGATTAAAAC	2940
	TTCTTCAACC GTAGTATCTT CAATTGATG CTTAACCCAT TCATGATCAA AAGTACTATA	3000
	CGCTAGCCAT CCACTATATA CACTCACAGC AATATCTCCG CATGAACTTA AACTTTGTAA	3060
15	CTTCATATTT GCAATCACTG CTAGTTTATA AATGTATAAA TTAGATAACT TCATATCATA	3120
	AAATTCATTT AATACTTTTA TAACTGACAC AAGTACTGCT GCACTTGAAC CTAATCCATA	3180
20	TTTATGACCA TTTGAATCAT CTAAATTACT ATCAATAGTC AGATGAAAAT GCTTCATCGC	3240
	TATATCGCAA CTTTTCGCGT ATTGTTCAAA TATTTCAATA GCTGTGACCA CATAATTTAA	3300
	TTGTTTTGCT GcATGTGGAT CTGAAATGAC AATACTATCT TCATCTCTAC TAAATGTAAC	3360
25	TGGGTTATGA TGTAATGCTT TTGAATGAAT GGTACCTTTA TATTGGTCTG CTTCTTCAAT	3420
	AGTAGCAGTT ACAAACGAT CTAACGCAAT AAGTACAGAT TTATATCCTG GTTCTGTTAC	3480
	AGCATATTCT CCAGCAATAT AAAGTTTTCC GGGTGCTTTG ACCTGAATCA TTTTATCTCT	3540
30	TCCTTACTCA ATTATTTCAA TTCCTGTGGC AATAATGTCA CTATCAATAA TTTGGTTATT	3600
	ATCAAAGTGT GTTAATAAAT TATCTATAAT CTGTTGCTTG TTTTCTTTT CTACAAGTAT	3660
	TTTCACATTA GGTCCCGCAT CCATTGTAAA ATAACACGGA TACCCCGCTT CTCGGCATT	3720
35	GTGAACAAGC GCCATGACAT CATACTTTT TGCACAAGA TATGTGAACG GCGGTGTTGA	3780
	TCCTAGATTG GTGGCATGCA TACGCAAACC ATTTTCTTCA ATTACTTAC CAAGGCGTTT	3840
	AAAATCTTTG TCTTGAATCG CTGCTTTTGC TTCAGCTAAA TCTTCATCAA TATGATCTAA	3900
40	CCAATATTGA TAAAACCTTG ATGTGTTTCG TGTCATGAC ATACCATATC GACTAGGTAC	3960
	CTTTTTAGAA TGTTGATTAA TCACAACAAA TATCATGGCA AGGTCATCTT CAAAATGATT	4020
45	CGATTCAAGT GGAACGGCAT ATGACGTCTC ATCACTATAC CCTTTTTCCC ATTCTGCAAA	4080
	TCCACCATAA ATACTACGCG ACGCAGAACC CGAACCAAT CGCGCCAATC TCGATAAATC	4140
	CTTATCTGAC AGCTGCATGT CTAGCGCTTG ATTACAAGCT GCTGCTAAAG CTGCATATGC	4200
50	GCTTGCCGAT GAAGCCAACC CTGCTGCTGT TGGTACAAAA TTGTCGCTTT CAATTTCTGC	4260
	ATACCAATCG ATGCCAGCTC TATTTCTGAC AATATCCATA TATTTTGAAA TTTTCTCTAA	4320

55

EP 0 786 519 A2

	AAAAGTGA	CTTCGTTTCAG	TGTAAAATTT	TTCTAATGTA	ACAGATATGC	TATTATTCAT	4440
	TGGAATGATT	AGTGCTTCAT	CTTTTTTACC	CCAATATTTT	ATAAGTGCAA	TATTCGTATG	4500
5	TGCACGTGCT	TTGCCACTTT	TAATCAACGC	ATTAACCTCC	TAAATTCTCA	ATCCAAGTAT	4560
	GTGCTGCACC	AGCTTTTTCT	ACAGCTTTTA	CAATATTTTT	CGCTGTTGGT	AAATCTTTGG	4620
	CAAGCAATAA	CATACTTCCA	CCACGACCAG	CGCCAGTAAG	TTTCCAGCA	ATCGCACCAT	4680
10	TTTCTTTACC	AATTTTCATT	AATTGTTCTA	TTTTATCATG	ACTAACTGTC	AACGCCTTTA	4740
	AATCCGCATG	ACATTCATTA	AAAATATCCG	CTAAGGCTTC	AAAGTTATGA	TGTTCAATCA	4800
	CATCACTCGC	ACGTAAAACT	AACTTACCGA	TATGTTTTAC	ATGTGACATG	TACTGAGGGT	4860
15	CCTCACAAAG	TTTATGAACA	TCTTCTACTG	CTTGTCTTGT	TGAACCTTTC	ACACCAGTAT	4920
	CTATAACAAC	CATATAGCCG	TCTAAACTTA	ACGTTTTCAA	CGTTTCAGCA	TGACCTTTTT	4980
20	GGAACCAAAC	TGGTTTGCCT	GATACAATCG	TTTGCATATC	AATACCACTT	GGTTTACCAT	5040
	GTGCAATTTG	CTCTGCCCAA	TTAGCCTTTT	CAATGAGTTC	TTCTTTCGTT	AATGATTTCC	5100
	CTAAAAATC	ATAACTTGCA	CGAACAAAAG	CAACCGCGAC	AGCTGCACTC	GATCCTAATC	5160
25	CACGTGATGG	TGGTAAATTC	GTTTGGATCG	TTACTGCTAG	CGGCTCTGTA	ATATTATTTA	5220
	ATTCTACAAA	ACGGTTCACC	AAAGACTTAA	GATGGTCAGG	CGCATCATAT	AACATACCAT	5280
	CGTAAACATC	GCTTTTAATA	GACGAATAGT	TCCGCTCTC	TAAGGCTTCT	ATTAAAACTT	5340
30	TGATTTTACC	TGCGTTAAAC	GGTACTGCAA	TAGCAGGCTC	TCCAAATGTA	ACAGCATGTT	5400
	CTCCTATTAA	AATAATCTTA	CCTGTCGATT	CCCCATATCC	TTTTCTTGTC	ATGTCAATAT	5460
	CACCTTTTAT	ATTTATCCTA	TACTTGATTG	ATTATTTTTA	TTTATTAGTA	AAAGACATCA	5520
35	TATTCTAAGT	TGCATngCAT	TCGCGTTAAA	TTTCATTGCA	GTCTTTATCT	CACATTATTC	5580
	ATAATTATGTA	TAATCTTTAT	TTTGAATTTA	TATTTGACTT	AACTTGATTA	GTATAAAACT	5640
	AACTTTCGTT	TACTTCAAAG	TTTAAATCTT	ATCGAGTGAT	ATTCAGATT	CTTTATCTTT	5700
40	TTATAAAATA	GCCCTACAAT	TTATAATTTT	CCACCCTAAC	TATAATACTA	CAAATAATAA	5760
	TTGGAATATA	TAGATTTACT	ACTAAAGTAT	TAGAACATTT	CAATAGAAGG	TCGTTTCTTT	5820
45	CATAGTCATA	CGCATTATAT	ATACCCTATT	CTCAATCTAT	TTAATACGTA	AAACATGAAA	5880
	TTTTCTTATT	AAATTTATTA	TTCCATCAT	ATCATTACTT	TTAATTTAAT	GATGTTCAAT	5940
	TTAAATATTA	GGTCAATAAC	ATATTTATGC	TTTTTATGGA	TACTTTCAAA	AATAACAGCC	6000
50	CCAAACGATA	ACTTGAAAGG	GGCTGTTAAA	TATTTAACTA	TTGCATTTGA	TCTATCATTT	6060
	TCTTGTTTCT	TTCAATCATT	TTATCAAAAT	ACGTATCGTA	TCTTTGCCAT	TCTTCTTGAG	6120

55

	TTACATCTTG AACAGTAATC GTTTTGTAA GCAATGTCTC TAATGAGGCC ATACAAGATG	6240
	GTTCAATTTT AGGATATTTA AATTTAGTCA CTTACCTTT TAAAGCATGT TCATAAAATG	6300
5	TTTGCATCAT CAATGCACGT TCTGAACCAG AGCCTTCAAC ACAAAGATAA ATTTGTACAG	6360
	CAATACCGCC TCTAACTCTT CGTTGCGATA TGCCTGCAAA TTTCTTACCA TCGATACTTA	6420
	AGTCAAATTT TCCTGGGCAA TAAGAATGTT CAATTTCCAT CGTATCAATA TCAACATTCT	6480
10	CATTTTCGAA CATTTTGCTA ATTAAGAGGT ACATCACAGT AAACGCTTCA TCAATCGTTG	6540
	TTTCTGTTTG TCCTTTGAAC ATCAGCGATA TATTTAATAC ACCTTGATCT AGAACGACAC	6600
	CTAAGCCACC AGAATTTCTA ACAATGGCAT TATAACCAAT CTCATTGCTT AAATAATCAA	6660
15	TGCCATCTTT TAAAAACGGC AATCTTGAAT CATGAATACC AAGAATAACA GTATGTTGAT	6720
	GAATCCAAGT ACGCACAAACA TTATCTGATA TATCTTTGCC CACACTTTTCG CAAAATGTAT	6780
	CATCGAATGC GAAAGATTGC ATAGGTTCTA ATCCAGAAGA ATGATCGATA TATCGCCAGT	6840
20	TGACGCCATT AAAATATTTA CTCGCTAAAT CCATCGTTAT TGTAAGGCTT GCGCTGCTGT	6900
	AATAATTGAA AGATTGTATA CATCTTCAAT TGAGCAGCCA CGTGATAAGT CATTTACTGG	6960
25	AGAATTTAAA CCTTGTAATA CTGGACCAAC TGCATCATAT CCACCTAAAC GTTGTGCAAT	7020
	TTTGTAACCA ATATTACCAG CTTCTAAACT TGGGAATACA AAGACATTG CATCACCTTG	7080
	TAATTTAGCA CCTGGCGCTT TTTTCTCAGC AACACCTGGT ACAATCGCAG CATCAAATTG	7140
30	GAATTCGCCA TCAATGATTG CTTCTAATTT TtCTTCTTCA GCTTTTtGTT GTGCTAATTT	7200
	GACAGCTTCT TGAACTTTTG TCACGTCGTC TGATTAGCA GACCCTTTTG TTGAAAAGCT	7260
	TAACATTGCA ACTTTTGGAT CCATGCCAAA GCTTAATGCT GATTTTGCAC TTTCTACTGC	7320
35	AATTTCTGCA AGTCCTTG TG AATCAAGTTC TGGATTGATT GCACAATCAC CAAAGATGTA	7380
	TTGTTCATCA CCTTTAATCA TAAAGAAGAT ACCTGATGTT CTTGATACAC CTGGTTTCGT	7440
	TTTGATGATT TGTAAGCTG GACGCACAGT GTCGCCTGTT GAATGTGCTG CACCACTAAC	7500
40	TAAACCATCT GCTTTACCAG CATAACAAG CATGTACCG AAGTAGTTCA CATTGTTTAA	7560
	TAATTCTTGT GCTTGTTCTT CAGTCGCTT ACCTTTACGT CGTTCAACAA ATGATTGAAC	7620
	TAATTCAGCT TTCAATTCAC TTGTCGCAGG ATTAATTAAT TCAATATTAG AAATATCAAG	7680
45	ATCAAGTTTT TGCGCTAAAG ATTGAACCTT AGTCTCATCA CCTAACACGA TTGGTGTAAAC	7740
	ATAATCTGTT GCTTGTAATT GTGTGCAGC TGTTAGAACA CGTTCGTCCT CTCCTTCAGG	7800
50	TAATACGATT TTAACGTTTT TACCAGAAAG TTTGTCTTTT AATACATTTA ATAAATCAGC	7860
	CATAATGTCC TCCTGTAATA TAAATCTTAT TAATCATTCA CGGTATAATT ATACGCCATT	7920

55



	TATGATAAAA	TTTATAAAGA	ACTGATGATT	TTTGAAAAGG	AGCGATAAAC	ATGAGTCAAG	8040
	CAGCCGAAAC	ATTAGATGGT	TGGTATAGTC	TACATTTATT	TTATGCAGTT	GATTGGGCAT	8100
5	CATTACGTAT	AGTTCCAAAG	GACGAACGCG	ATGCACTTGT	CACTGAATTT	CAATCATTTT	8160
	TAGAAAATAC	AGCAACTGTA	AGATCATCAA	AATCTGGTGA	TCAAGCTATT	TATAATATAA	8220
	CTGGTCAAAA	AGCAGATTTG	TTATTATGGT	TCTTACGTCC	TGAAATGAAG	TCTTTAAATC	8280
10	ATATTGAAAA	TGAATTTAAC	AAATTGCGCA	TTGCTGACTT	CCTAATCCCT	ACATATTCAT	8340
	ATGTATCAGT	CATTGAATTG	AGCAATTATT	TAGCTGGTAA	ATCTGATGAA	GATCCTTATG	8400
	AGAACCCTCA	TATCAAAGCA	AGATTATAACC	CAGAATTACC	ACATTCTGAT	TATATTTGTT	8460
15	TCTATCCAAT	GAACAAACGT	CGTAATGAAA	CTTATAACTG	GTACATGTTA	ACTATGGAAG	8520
	AACGCCAAAA	ATTAATGTAT	GACCATGGTA	TGATTGGTAG	AAAATATGCT	GGCAAAATCA	8580
20	AACAATTTAT	TACTGGTTCT	GTAGGGTTTG	ATGATTTCGA	ATGGGGCGTA	ACATTGTTCT	8640
	CAGATGACGT	ATTACAATTC	AAAAAAATTG	TATACGAAAT	GCGCTTTGAT	GAAACAACAG	8700
	CACGATACGG	TGAATTCGGT	AGTTTCTTTG	TAGGACATAT	TATTAACACA	AACGAATTCTG	8760
25	ATCAATTCTT	TGCGATTTCT	TAATACATTG	GTACGTTTAT	AAATTAATAA	AAAAATTCCA	8820
	AGCTTATCGG	TTTAAGCTTG	GAATTTTTTCG	TTTATCTTCA	GTATATTCCC	GTATACATAA	8880
	GACGTGATTT	GGTAAATAGT	TGAAATCTGT	ATGTTTAAAC	TTATATATAT	GTGCTAATGT	8940
30	ATTATCAATA	ACAAAGtACA	CTTGCTCAT	AGCAAgTsac	CCGAgTAGTC	TTCCTTGGGA	9000
	GAACTTTAAC	TACTATCACT	ACATATAAAC	GTTAACCTCA	ATAGAAATTA	TACAGTCGCT	9060
	ACTCTATACA	ATTTTGTAA	TGGTTAACTA	ATATTATTTT	AACCTATTTG	AAATATTTGA	9120
35	AACATATTTT	TGTCGAATTT	TTTTCAATAA	TTTTTCCTTT	TTATACTTCA	AGAGAATTTT	9180
	AAC <del>T</del> ACTAAA	AATTCCGATG	ATTATTATTA	CAATAGTATC	AAATATTAGT	TTTTTAAAAT	9240
	CAATAACAAC	TTATCAAAAA	GCTCATGTGG	TTATTTTATA	GTGTATAAAC	TATAATGAGT	9300
40	ATTAAATTCT	TATAAACAAT	GGTGATGAAA	TGGACATAAA	TTCAGAAGAA	TACAAACAAG	9360
	AGGTACTTAT	CAAAGACGTT	GTCATGCTTG	CTGCTCGCAT	ACTATTAGAA	TCTGGTGCAG	9420
	AAGGTACGCG	TGTAGAAGAT	ACCATGACAC	GTATTGCAAA	AAAACCTGGT	TACAGTGAAA	9480
45	GTAACAGCTT	TGTTACAAAC	ACTGTCATCC	AGTTTACGTT	ACATTCGGAA	TCGTTTCCTA	9540
	GAATATTTAG	AATTACCTCT	CGAGATACAA	ACTTAATAAA	AATTTCTCAA	GCTAATAAAA	9600
50	TTTCGCGTCA	AATTACAAAC	AATGAAATTT	CTTTAGCCGA	AGCAAAAACG	CAACTTGAAA	9660
	AAATATATGT	TGCTAAGCGT	GACAGCAGTC	TTCCCTTTAA	AGGTTTGTGCT	GCAGCAATGA	9720

55

TAGCAGGTAG TCTAGGATAC CTAGTCACTG AGATTTTAGA TCGTAAGTWA CACGCACAGT 9840  
 TTATCCCAGA ATTCaTTGGT TCaTTAGTTA tTGGGATTAT CGCCGTTATT GGACATACAC 9900  
 5 TTATTCCAAC AGGTGACTTG GCAACTATTA TCATTGCGGC AGTCATGCCT ATTGTTCCCTG 9960  
 GTGTATTAAT AACAAACGCA ATACAAGATT TATTTGGTGG ACACATGTTG ATGTTACACAA 10020  
 CGAAATCATT AGAAGCATTG GTTTnGCGTT TGGCATCGGT GCTGGCGTTG GTAGCGTATT 10080  
 10 AATTTTAGTA TAGGAGTATC AGACTATGTT TTGGATCTTA AACTTTATCT TTAGCTTTTT 10140  
 AGCTTC 10146

(2) INFORMATION FOR SEQ ID NO: 244:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2022 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

25 ATTTAATTGG TTGGTGGCGT ATTCTChTTT CCAAGATTG CCACGATTGG ATGTTTCAGTC 60  
 GACAAATATG TAACCAGTCC CGATAAATTA CGTTCCTCAA TCATTGATAG TATGTCTGGT 120  
 GTTTGTAACG ATTTATCGAG TAAGGCATCA AGCAATTGAT AATGTCCCAA TACAACAAAA 180  
 30 TGCACGTTGT CTCTCAACTG CTGTTGAATA AACTGAATAA AGAGCTTTAA GCTCTGTTGC 240  
 ACATTCGCTA ATGATGGTGC ATAGTTTTCC AAACCAACTT GTACAGCCGC TTCATTATTC 300  
 CGAATGATTA AACCTGTGTA TGCCACTTTT GTTGCTGCAG TTGGATACAT TGAGTAATAA 360  
 35 CGCAATAATT GATCTGTAAA ATCATTTCGA AGTGCATAAA TTTGATGCTC ATGTTGCCAA 420  
 AAATtTCGCT CACCCATCTG CTgCAAATCC TCATGGTTCA ATTGTTTCCA GTCCAACTTT 480  
 TCAAĊCACAC TAAAATCAAC TAACTCATAA TCCGCTTTAT TAAAATATTT TAAAATGCT 540  
 40 GTTTCGATT CTTTTAACGC AATTAATTGT TCTGnATTAT TCACTCGACC ACCCTTTACT 600  
 TTCAATACTG TATTTAAAT CACTTGGTAT TTTCTGTGTT TGCTTTACTT CTCTACCACG 660  
 CTAAAGTGTA ATATGATTAA TAACTTATCA TTTTtagCAA TACATTACAA CCTTTTTTCAG 720  
 45 AAAATTCGGT GTATTGATTT TAAAATTTTT TAAAATanAA AAGGCAAGAC ATTTGTGCCT 780  
 ATAAAAATGC TTAACCAAGA TTTTtATATT GaAGTTGTAC TTCTTGACA TATTGTCCTT 840  
 50 GCCTTATTAT GTAAAGTTAT TTTCTTTCTA TCTTTTIATT AAATTTAACT ATTCTTCATA 900  
 ATCCCGATT CTTTAAAGT AACGTCTATC TTGTTTACTA TATACATTTT CAGGATTAAA 960

55

TTGGTAAACG TTCGTTGCTG ATATATCTGT AAAATTGTTT GGACCGACAC CTGCAATAAA 1080  
 CTTAAACTCT GCTTCATCTA CCAAATAATC ATACGCTTGT GTATGTCTAT CCTGTGCGCC 1140  
 5 ATGTGGAAAT ACAAACATAT CTGTTTTACC TACAATTGGT TCAACTTCAT CTTTCCATCT 1200  
 TTTAGTATCA CGTTTAATAC CTTCTAAAGA TGTTTTTTCA AAATTAATGT GACCATATGA 1260  
 ATGACTCGCA AATGACCATC CATCCGTTT CATTGCGCGA ACAACTTCCT CAGCTGCCTT 1320  
 10 TTTATTCTTT GTATAATCTT TACTCGTTAA TTCATTCTGT CGATAACCTA ATACGCCCTC 1380  
 ATAACCGGTT AAAGCAACAA CACCTTTTTT ACCATTTAAA GAAAAATCTG GATGCTCTTT 1440  
 TACAAATTTA TTAAAATTG GCACGATATC ATTGTCATCA GAATAAGTAG CATGGCCTTT 1500  
 TTTGTCTGTA GTTTCAGAAA CAACATGTTT ATTTTATCG AGTACTAAAC GGTCAGCATA 1560  
 ACCATGGTGT CTCATGTAAC TATAGTAATT CATATCATCA ATTGAGATGA TTAGTGGCTT 1620  
 20 TTTACCTTTC GGCAATTTTA TTTTTTTGGC TTTTACATGA TGAGATGATA AGTCGTATAC 1680  
 ATCATGTGGA TTAACGATGA TGTAATTATT TTTATATAAT TCGTTCAATG ATTTTTTAAA 1740  
 TTCACTTACA GTAATCATCC AATCATTGTT GCCCTTAGCT TGGTGTGTAT CTCCTGTAAA 1800  
 25 CGCAACTTTT GGGTCTGTAA TTAATGGGTG ATAAACACA TGATAAACTT GGCCGTGATA 1860  
 TGTTCCTCAA TGTTTCATCCA TTTTcGATT aTGCTTTGCA TACTCATTG GATTAACAGA 1920  
 TTTATTkTGA GCTTCTCAT TTGCTTGGA ACAGCTATAT mACAATGCAA CTGATAATAA 1980  
 30 CAGAAAAAAT AGCAATAAAT ATTTTTTATG CATTAAACAT TC 2022

## (2) INFORMATION FOR SEQ ID NO: 245:

(i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 1340 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

ATAGAGTGAT AAAGGATGGT TGTCATATGA TAAATGCAGT AGTAATAGCA GTAATTTTAA 60  
 45 TGATTGTGCT ATGTTTATGT CGATTAAACG TnTTATTAAG CTTATTTATC AGTGCCTAG 120  
 TTGGTGGCTT AATTTCAGGC ATGAGCATTG AAAAAGTTAT AAATGTATTT GGGAAAAATA 180  
 TAGTCGATGG TGCTGAGGTA GCATTAAGCT ATGCTTTATT AGGTGGATTT GCAGCATTAA 240  
 50 TTTCATACAG TGGTATCACA GACTATTTAG TAGGAAAAAT TATAAATGCA ATTCACGCTG 300  
 AAAATAGTCG ATGGTCAAGA GTTAAAGTCA AAGTGACAAT AATCATTGCA TTATTAGCTA 360

55

CACCATTGTT AAGTCTGTTT AATGACTTAA AAATAGATAG ACGTTTAATC GGTTTGATTA 480  
 TCGGTTTTGG TTTATGTTTC CCGTATGTGT TATTACCATA TGGATTCCGT CAAATTTTCC 540  
 5 AGCAAATTAT TCAAAGTGGC TTTGCAAAGG CAAATCACCC AATTGAGTTT AATATGATTT 600  
 GGAAAGCAAT GCTTATTCCT TCAATGGGGT ATATTGTTGG CTTACTTATC GGTTTATATG 660  
 TATATCGTAA ACCACGTGAA TATGAAACAC GTAAAATTTT AGATAGTGAC AATGTTACAG 720  
 10 AGTTAAAACC ATATATCTTA ATAGTAACAA TTGTAGCAAT ACTAGCTACA TTTTATAGTAC 780  
 AAACATTTAC AGATTCAATG ATTTTGGTG CACTGGCAGG GGTACTCGTA TTCTTTATTT 840  
 CACGTGCATA TAATTGGTAT GAATTAGATG CTAAGTTTGT TGAAGGTATT AAAATTATGG 900  
 15 CTTATATTGG TGTAGTTATT TTAACAGCAA ATGGATTTGC TGGTGTAAAT AATGCTACTG 960  
 GTGATATAGA TGAATTAGTT AAACTTTTAA CAAGTATTAC TGGTGATAAT AAATTATTTA 1020  
 GCATTATCAT GATGTATGTG ATAGGTTTAA TTGTCACTTT AGGTATTGGA TCATCATTG 1080  
 20 CAACAATTCC TATTATCGCA TCATTATTCA TTCCTTTTGG AGCGTCAATT GGA CTAGATA 1140  
 CAATGGCATT AATCGCATTG ATTGGAACAG CGAGTGCATT AGGTGACTCA GGTCGCCTG 1200  
 25 CAAGTGATTC AACATTAGGA CCAACTGCGG GATTAAATGT TGATGGCCAm CATGATCATA 1260  
 TACGTGATAC ATGTGTACCA AACTTCTTGT TTTATAATAT TCCTTTAAAT GATTTTCGGT 1320  
 ACTATTGCTG CTATGGTACT 1340

30 (2) INFORMATION FOR SEQ ID NO: 246:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3365 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

40 CAAAATCTAA GAGAATAAaA TTTGTTAATT TnAAATAGCA AGCAATTCAA AGTTATATGT 60  
 GTAATAGATA AAATAGATAT CCCTATAGTG ATGCGTTACT AGCTAAACAT AATAACACAT 120  
 45 TAGAAGATAA TGAAGTTAAG GAGTTACTGG ATTGTTTCGA CTATGTAAAT AAGTATAAAA 180  
 ATATCCAACG ACAAACGTA ATTATAAAAT GGTAAAAGCT ATGGTACAGT TTCAAATTGC 240  
 TAATGACATG CGTATCGGTG AGCTACTTGC AATAAAGAGA GTAAATATAA ACTATGAAGA 300  
 50 TAAAACGCTA GATATCGACG GTAAAGTTAA TTGGATAACT GAAAAAAGAC GGGAGCATTC 360  
 GGAGTAAAGG AGACAACCTGA AAGAAGTAAT AGCTATAAGG CCACAGGGCT CACTACCCAA 420

55

	TATTCACAAA TACGTCTGGT AGCCCTATCG ACTCGAACAA AATTAGCCAC ATTATTAAAG	540
	GGGGGCGCTG ATATTAGTTC TATTAAGAAA CCTATAACGA CGCATACATT ACATCATTCG	600
5	CATATATCTA CACTTGCTCA ATTAGGAATT AACTTAAAAG CAATGCAAGA GCATGTAGGT	660
	CATTGAGATT ATAAAArAAA TCTAGAGATA TACACACATG TTAATAATCA GATGGCGAAA	720
	GATATGATGA ATAAATTTGA ACGATTGGGG AGTTAAAATT GGAAAAAGAT GaTACACTAG	780
10	CAGAAATTAA GCCTATGCTC AATTTTGTATG AGCAAATAGC AAAATTAAAA CAGATGAATA	840
	TATTTTTTAA TATTATTGAC ACCGAAAAAG CAAATGAAAT TCTTAGAAAA AATAATTACT	900
15	TCTTCAAAC T wGcTTATTTTC CGaAAAAATT TCGrAAAAAA GaATGGCGGC TATTTTCATAG	960
	AATTTGCTTA TTTATCAGAT TTAGCAACTA TAGATATGaA ATTAAGATAC ACAATGTTGC	1020
	ATTTAACTTT AGATATTGAA CATAGTTTAA AGTATCTAGT CTTAAACTA ATAACAGAAA	1080
20	ATAACCAAGA AGATGGTTAT AAAATAATAG ATGAGTTCTT ATGTATTGaT AAATCATATA	1140
	GCAATTCAAA TTTTGACACA AATTCAAGAA CACCAGAAGA AGTTATGGAA ACCAAAATCA	1200
	AAAATAAAAA CGAAATATTC AAGCATATGA ATAAACGAGG ACAACTACCC GAGAAGTTGa	1260
25	ATAAATACTA TcmAAATCCA CCCGcnnGGk TTTGCaTTGr ATTCATGCAA CTAGGTCAAT	1320
	TCGTTTCGTT TCTCAACTTC TATTACAAGA AGTACAATGA CGAAGAATTG AGAGTTGCTA	1380
	ATATTTTAAAT GCCTTTAGTT AAAAATATAA GAAaCAAATC AGCTCATAAC CAACCCATCA	1440
30	TAGCAAATCT AAATTATGAC AGTAGATTAC CTCAATATTT ATTTGAAAAA GGAATAATA	1500
	TAGGCATATC TAGAAACATG TTCGGAATAA AAAATTTTCAT AGATACTkTC kSTACGCTAG	1560
	AATTACATAA TCAAGTTTGT AGTAATGCAA TTATCCAAGC AAGATATCAC GATTGGACC	1620
35	AACCTCAAAA GCGATATAAA AGrAACGrAA GCTATTATAA TAATGCATTA GCTATCAAAA	1680
	GATTfTTTAT AGCTTTAGAT AAAATTATTG ACTTCAACAG ACCAAAAGTA TAACTATCT	1740
	AGTGAGGAAA GAGACTTATA GGTCTCGCGA GTTATTTTAA TTCGTATGCA AGAAAAAGAA	1800
40	GAGCTATGCA TTTTATTTAA AATGCGTAGT TCTTtTTTTA TGCATCTAAA TTCATATTAT	1860
	TTTTGCAATA TAAACATATC TTTGTGCAAA TTCCGAACAC AAAACATTCA CATCATCCTT	1920
45	TTTtGCCCTT TTTCTATACC CCAAAACACA AAAAGCCCCG TAAGCCTATG CcTACGGGgT	1980
	TTGACAATAA ATTATATATT ATTGTTCTTC TTAAACATAT GGTAATAATG CCATATGACG	2040
	AGAACGTTTG ATAGCTGTAG TCAaCATACG TTGATATTTA GCTGAAGTAC CAGTTACACG	2100
50	ACGTGGTAAA ATTTTACCGC GTTCTGAGAT AAAACGTTTT AATAATTCAG TGTCTTTGTA	2160
	GTCGATATGT GTAATACCAT TTGCTGTGAA ATAGCATACT TTTTACGAC GACGTCCGCC	2220

55

EP 0 786 519 A2

CGTTAATTTT TATTAGAATG GTAAGTCATC ATCACTTATA TCAATCGGTC CGTTTGCATT 2340  
 TGCAAATGGA TTATCAGATT GTTTCGTGTT TGATGAATTA TTGTACGAAT TGTTTTGTCC 2400  
 5 TGATTGTTGA CCACCGAATC CTTGACCGTA ATCTTGGAAT TCATTTTGTG GACGTTGGCC 2460  
 ACCATTTTGT TGCGCATTTT TAGGTTCAAG GAATTGAACG CTATCACACA CAACTTCAGT 2520  
 AACAAACACA CGACGACCTT CTTGATTTTC ATAATTACGG GATTGTAAGC GACCATCTAC 2580  
 10 ACCAGCTAAA CTACCTTTAG ATAAATAGTT ATTTACATTA TCTGCTTGTC TTCTAAAAAC 2640  
 AACACAGTTA ATAAAATctG ctTCGCGCTC CCCTTGAGCA TTCGTGAACG TACGATTTAC 2700  
 TGCAAGAGTG AATGtCGCTA CACTCACACC TGAGGGAGTG GTTCTGTATT CCGGATCTTT 2760  
 15 CGTTAAACGA CCTACTAATA CAACTCTATT TAGCATTTAA ACGCCCCCTC TAATTATTAC 2820  
 TTGTCTTCGT CTTACGAAT AACCATGTAA CGAATGATAT CGTCACTGAT TTTAGCTAGA 2880  
 CGTTGGAATT CGTCAGTAGC TTTGTTGTTA TCAGATTTAA CACGTACGAT GTTGTAGAAG 2940  
 20 CCATCTTTGA AATCATTGAT TTCATAAGCT AGGCGACGTT TACCCAGTC TTTTGCTTCT 3000  
 AAAACTTCTG CACCTTCAGT AGCTAAGATA CCGTTGAAAC GTTCAACTAA CGCTTTTTTA 3060  
 GCATCTTCCT CAATGTTTGG GCGTACGATG TACATAACTT CATATGTTCT CATTTTATAT 3120  
 TTGCACCTCC TTGTGGTCTA TACGGCTTAT CAATCTTAAA ACAGATAAGC AAGGaATAAT 3180  
 TTTCACTACT CACAATAAAG AATTaTATCA TGCGCCATTA CTTTTTACAA TaATAATTcA 3240  
 30 AACTACTCTT CATATCATT TTAGATAtAA TTCATTTGaA ACTTTcNaTG ATATTTTnAA 3300  
 AAATACACTT CACAAAAGCG AACATATGTn CTATAAnAGT TGTGAGGTGG TAAGGAATGA 3360  
 ATTTA 3365

35 (2) INFORMATION FOR SEQ ID NO: 247:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1032 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

GTTAAAAGTA ATTGGTGGTA TTGATGATGA TTTTACAGCC AATGTTATGC ATCCAAATCA 60  
 ATATCGAATT CGATATTCGT CTCAGAAACA GGACCTTAAT GAAGATATGA CAGTTTTTGA 120  
 50 TGCAGTATTA AGTTCTGATA CAACAACTTT ACGCATCATC AAGCAATATG AGCAGGCAGT 180  
 ACAAGCTTAT GCGGATGACC AAAGTGATAA ATTGTTCAAG CGAATGATGG ATGCGCAAGA 240

55

ACTAGGTATA CATGATACTA CTAAATACAT TAAAGAATTA TCCGGCGGAC AACAAAAACG 360  
 TGTGTACTT GCTAAACAT TAATAGAACA ACCAGATTTA TTGTTATTAG ATGAACCTAC 420  
 5 GAACCATTTA GACTTCGAAT CAATCAGCTG GTTGATCAAT TATGTGAAGC AATATCCTCA 480  
 TACTGTTTTA TTCGTAACCC ATGATCGATA TTTTTTAAAT GAAGTTTCCA CTAGAATTAT 540  
 TGAACTAAAC AGAGGTAAGT TAGCGTCATA TCCTGGTAAC TATGAATCTT ATATTGAAAT 600  
 10 GCGCGCTGAA AGAGAAGTAA CACTTCAAAA GCAACAACAA AAGCAACGAG CTTTATATAA 660  
 GGAAGAACTT GCTTGGATGA GGGCTGGgAG CTAaggCTCG TACTACAAAG CAACAAGCTA 720  
 GAATTAATCG ATTTAATGAC CTAGAmAATG AAGTTaACCA GCAATATAAA GACGATAAAG 780  
 15 GTGAATTGAA TCTTGCTTAT TCaAGATTAG GTAAGCAAGT GTTCGAATTA GAAGACTTAT 840  
 CAAAGGCTAT TAATGATAAA GTATTATTTG AACATCTGAC GGAAATTATT CAAAAmGGTG 900  
 AGCGTATTGG TGTGTTGGG CCAAAATGGAG CTGGTAAAC AACACTCTTA AATATTTTGA 960  
 20 GTGGAGAAGA CCAACAATTC GAAGGTAAAT TGAAGACTGG GCAGACGGTT AAAGTAGCTT 1020  
 ATTTTAAGCA AA 1032

25 (2) INFORMATION FOR SEQ ID NO: 248:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 852 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

35 TGTGATTAAAC GAAGCTTATT TTCGTACACC TTCAACRACT GATTACAACG GCGTTTATCA 60  
 AGGTATTAT ATTGATTTTG AAGCAAAGGA AACTAAAAAC AAGACGTCCT TTCCTTTAAA 120  
 TAATATTCAT GACCATCAAG TCGAACATAT GAAAAATGCA TATCAACAAA AAGGTATTGT 180  
 40 GTTTTTAATG ATTCGTTTTA AAACGCTAGA TGAAGTTTAT CTTTTACCCT ATTCAAATT 240  
 CGAAGTATTT TGGAAGAGAT ATAAAGATAA TATTA AAAAG TCTATAACAG TTGATGAAAT 300  
 ACGAAAAAAT GGTTACCATA TTCCTTATCA GTATCAACCA AGATTAGACT ATCTAAAAGC 360  
 45 AGTTGATAAG TTGATATTAG ATGAAAGTGA GGACCGCGTA TGACGGAAAA CAAAGGATCT 420  
 TCTCAGCCTA AGAAAAACGG TAATAATGGT GGGAAATCCA ACTCAAAAAA GAATAGAAAT 480  
 50 GTGAAGAGAA CGATTATTAA GATTATTGGC TTCATGATTA TTGCATTTT CGTTGTTCTT 540  
 TTACTAGGTA TCTTATTGTT TGCTTATTAT GCTTGAAAG CACCTGCTTT TACCGAAGCT 600

55

TTAGATAATG GCCAAAGACA TGAGCATGTA AATTTAAAAG ACGTGCCGAA ATCAATGAAA 720  
 GACGCAGTAC TTGCAACTGA AGACAATCGT TTCTACGAAC ATGGCGCACT TGATTATAAA 780  
 5 CGTTTATTCG GTGCAATTGG TAAGAACTTG ACTGGTGGAT TTGGkTcGA AGGtGCCTCA 840  
 ACATTAACAC AA 852

(2) INFORMATION FOR SEQ ID NO: 249:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5804 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

20 CACTTTTTTC ATTAAAAATC TCATATTTAT AACTGAAACG TAATCTCGAA TATTTTTCAA 60  
 CCCAAGTTTT AACTTTAACT TTTCTGGAT AAAAAATAGA CTTTTTATAA TTGACATTGA 120  
 GGTCAGTCAC AGGTGAAATG ATTCCTTGTT TTTCCATATC AGCATAACTA AAACCTAACT 180  
 25 TCGATATATA ATCCAACCGC GCAACTTCAA ACCAAGTTGC ATAATTCCCG TGATAAATTA 240  
 CACCCATCTT ATCAGTTTCA GCATAACGCG CTCTATTTC TGTAATACTA TATATCATTT 300  
 TAAGCCTTCT TTCAGTTTAA CTTTATATCT CATTCTAACA TAAAATACAA GAAGAGGCCG 360  
 30 GCCAAGAACA CAAAGGkTTT GAACCGACCT ATTATATCAT AAaGTTTATA GAAGTATTTT 420  
 TGAGCACTAT CAAAGTGCCT CAAATACCGA TTAAAATTTT ACTGTGATAT CTATTTTTTA 480  
 TTGCGCTAAT TTATTTCTTA AAACCATTG TAAAATTCCA CCGTGACGAT AGTAATCCAT 540  
 35 TTCAACAAGT GAGTCAAAAC GAACCATAGC GTCAAATTCT ACCAAATCAC CATCTTGCTT 600  
 CTTAGCAGTA ACTTTGACGT AGtCATGTGG TTGAACATTT TCATCAATAT TAACAGrAAT 660  
 TTCTTCTGTA CCATCTAGAC CAAGAGAATC AGCTGATTCA CCTTTTTTAA ACTCTAATGG 720  
 40 TAATACACCC ATCATAACTA AATTGAAACG ATGGATACGT TCATAACTTT GTGCAATAAC 780  
 TGTTTTAACA CCTAATAAGT TTGTACCTTT TGCTGCCAG TCACGAGATG AACCCATACC 840  
 45 ATAATCGTTA CCAGCTAATA CAACTAAACC TGTACCATCT TCTTTATATT TCATTGCAGC 900  
 ATCAAAGATA GGCATTACTT CATTTGTTGG CCAATAAGTT GTAAAACCAC CTTCAGTACC 960  
 TGGCGCTAAT TGGTTTTTAA TACGTATATT AGCAAACGTA CCTCGAACCA TTACTTCGTG 1020  
 50 ATTACCACGT CTTGAACCAT ATGAATTAAA TTCACGAATA GGCACCTGAT GATCTTGTA 1080  
 ATATTTACCA GCTGGCGTAT CTTTACCAAT TGCACCTGCT GGAGAGATGT GGTCAGTTGT 1140

55



	TTCTTTAGAT AATCCTTGGA AGAATGATGG ATTTTGAATG TATGTTGAAT TAGGATCAAA	1260
	GTCATATAGA GGTGATCAG TTACATCAAT CTCATTCCAT AATTCGTTGT TATTGTATAC	1320
5	GTTATTATAT TCTTCAATAA ATAATTCAGG TGTTACAACA CTATCAACGG TATCTGAAAC	1380
	TTCTTTAATT GATGGCCAAA TATCTTTCAA ATATACATCT TCACCGTCAT TACCTTTACC	1440
	AATAGGTTCA TTTTGTAAT CAATATCAAC CGTTCCAGCT AATGCATAAG CAACAACATA	1500
10	CTGTGGTGAA GCTAGGTAAT TGGCTTTAAC AAGAGGATGG ATACGACCTT CAAAGTTACG	1560
	GTTACCAGAT AATACAGATG TCACTAATAG GTCCTCATCA GCAATCGCTT TTTCAATTTC	1620
15	TGGTAATAAA GGACCTGAAT TACCGATACA AGTTGTACAT CCATAACCAA CCAAGTTGAA	1680
	GCCTAAATCA TCTAAATAAG GTTGTAGCC AGCATCTCTT AAATATCCGG TAACAACCTT	1740
	TGATCCTGGT GCTAGAGAAG TTTTAACGTA TTCAGGAACT TTCAAGCCTT TTTCAACTGC	1800
20	TTTTTTAGCA ACTAAACCTG CACCTAACAT TACATAAGGG TTAGATGTAT TTGTACATGA	1860
	TGTAATTGCT GCTATTGCAA TATCACCTGT TTTTATTGTA GCTTTTGATC CATCTTTAAA	1920
	GTTAATTTCA GCTTTCTTAT CAAATTCACT TTTATCTAAA CCGTGTCTT GGTGCTGC	1980
25	TGGAGCTGTT ACAGAAATTT CAAATGATGA TTTTATATCA CTTAAGAAAA TTAAATCTTG	2040
	AGGACGTTTT GGTCTGAAA GCGATGCTTC AACTGTTGAT AAATCCAATT CGATAACATC	2100
	TGTATAATTA GGATCTTCTT TCTCAACATC AAAGAACATA TGGTTTTGTT TCAAATATTC	2160
30	TTTTACTAGC GCGATATGTT CGTCTGATCT ACCAGTTAAC TTCATATATT TAAGAGATTC	2220
	ATCATCAACT GGAAGAATC CGCAAGTTGC TCCATACTCT GGTGCCATGT TTGCAATTGT	2280
	AGCACGGTCT GCTAGTGGTA AATGTTGTAC ACCTGGACCA AAGAACTCCA CAAATTTACC	2340
35	AACAACACCT TTTTACGTA GCTCTTGAGT TACTCTTAAC GCTAAATCAG TTGCTGTTGC	2400
	GCCTTGTTGGT AATGAATTTA CTAGTCGTAC ACCAATAACC TCTGGAATTG GGAAATAAGA	2460
40	AGGTTGTCCA AGCATTCCAG CTTTCTGCTT AATACCACCA ACACCCCATC CTAGTACGCC	2520
	AATACCATTT ATCATTGTTG TATGTGAATC AGTACCAACT AATGTATCTG GAAATGCAGT	2580
	TTTTTCACCA TCTACATCAC GAACATGTAC AACACTTGCT AAATATTCTA AGTTAACTTG	2640
45	GTGAACTATT CCAGTTGCAG GAGGAACTGC ATTGTAATTA TCAAATGCTT TCGTTGCCCA	2700
	ATTTAAAAAC TGATAACGTT CATAGTTACG TTCAAATCTT AATTTTCATAT TACGTTCAAG	2760
	AGCTTCTGGA TTTGCATAGC TATCCACTTG AACTGAGTGG TCAATAACTA AATCCACCGG	2820
50	TACTTCTGGA TTAATTTTAG TAATATCTCC CCCAACGTCA TCCATTGCTT TACGTAAAGA	2880
	AGCTAAATCA ACTACGGCTG GTACACCTGT GAAATCTTGT AAAATAACAC GAGAAGGTTT	2940

55

GTCTGTAATT ACAAATCAT CTTCTTGACG AAGTAAAGAT TCTAACAAAA CACGAATTGA 3060  
 ATAAGGTAAA TTGGAACTT TAGTAATACC TTGCTCTTCT ACAGCTTTTA AATCATAGTA 3120  
 5 AGTATAACTT TGGCCATTCA AGTCAAAATG TTTTTTTGAT TGCTCTTTAA AATTGTCAGC 3180  
 CATTTAATGA TCCCCCTGA TACATTTTTA TATTTATATG CCTTGATTAA ATTGTATTAT 3240  
 TATATTTATT GATAACAAC TCATCATGCT TAGAAAACGC TTAATTTAGG TTTTGACTTT 3300  
 10 TTAATCAGAG TATATAAGCA AACTTATCA TACAGGTAAG GTGTAATAAG TATTTTTTAT 3360  
 TAATTGAGAA TAATTATCAA TTTTCGCAAT GATTCAATTC AATTTTTTAA CGTATTATTT 3420  
 CATTGAGCAG AAAGAAAATT ATGGCACCAA ACTTTAATAT TTTTTTCAAT GTCATTCTTT 3480  
 15 TGATGGGAGT GGGACAGAAA TGATATTTTC GCAAAATTTA TTTTCGTCGTC CCACCCCAAC 3540  
 TTGCATTGTC TGTAGAAATT GGAATCCAA TTTCTCTTTG TTGGGGCCCA TCCCCAACTT 3600  
 GCACATTATT GTAAGCTGAC TTTTCGTCAG CTTCTGTGTT GGGGCCCTCA CCCCAACTCG 3660  
 20 CATTGCCTGT AGAATTTCTT TTCGAAATTC TCTGTGTTGG GGCCCTGAC TAGAATTGAA 3720  
 AAAAGCTTGT TACAAGCGCA TTTTCGTTCA GTCAACTACT GCCAATATAA CTTCGTAGAG 3780  
 CATAGAATAT TGATTTATGT CCCAGCCTGA GTTAATTTTC TATAAAAGTA TATTTAATTT 3840  
 25 GCGTTTATAC CGTCAAACTT CACTTTAGCT TTGTCAAACC CCTTTCTATT AAGTTTTCAG 3900  
 AAATAAACCT ATCTTAAAT ATAAAAAAT CGAGAATTCG TAGTTTAATA ACGAAATCTT 3960  
 CGTTCTTATC CTTTTGAATA TACTCAATTT TCCACAAAAA CAAACAAGTA GTATATCTGT 4020  
 TCTAGCTACT AGAATGACAT ACTACTGTT ATTAAATAC TTAATAAAC TTTATTAGTT 4080  
 ATCTTTTTTC TCTATATTC TACGTGACTG ACGCTTTTCA AGAATGTCAG ATTCATAATC 4140  
 35 TTCTTGTTGA CTCTTGATAT ATTCTTGTA GCGATGTTTA TTCGGAGTCA ATGTTAAACC 4200  
 TAGGAATTTA CGTTCCTGGT TCGCATCCTT GTAGAAACTT ACCATCATGA GTATGACGAC 4260  
 AAAGGAGAAT GGAATGCAC TTATAATTGC AGCACTTTGA ATCGCATTTA AAGCTTCAGC 4320  
 40 GCCGTTACCG CCACCAGCTA ATAAAAGTAC AAATGCTATT AAGGCCTGTG AAATCCCCCA 4380  
 AACAACTTTT ACCATACTAG ATGGATTAA TGAACCAAAT GTTGTTTGCA TTCCTAATAC 4440  
 AAATGTTGCT GAGTCAGCAG ATGTAATAAA GAATGATGCA ATTAATAATA ATGCAATCAA 4500  
 45 CGATAAAACA ATGCCAAATG GCACATGATT AACACTCCA AATAGCTGTG TTTCAGGAGT 4560  
 CATATCAAAA ATTTCTTTGT GTTTCTTACC TGTCTCGATG CCTAATACAC CAAAGACACT 4620  
 50 AAACCAACA AACTAACA TGTCTGGAAC TAGCAAGACA CCAGAAATGA ACTCTCTAAT 4680  
 TGAACGTCCT TTTGAACTC GTGCAATAAA CACTCCAACG AATGGACTCC AACTTAACCA 4740

55

TGCTGTATCA AAAC TATTAA ACAAGAATGT GTTTAGTAAA CTACCCGTAG AGCTAGTTAA 4860  
 CATATTTAAA ATAAGAACAG TTGGTCCAAC AATTAAAGCA GCTACCATTA AAATAGTACC 4920  
 5 TAAACCAATG TTCAAGTTAC TTAAGTATTG AATACCTTTA CTTAATCCAG ACCATGCACT 4980  
 TGCTATAAAT AAGATAGTAA CAACAATGAT GATAATCGCT TGTACAAACG TATTGTTTGG 5040  
 AACATTGAAC AAGTAATGTA AACCACCATT AATTTGTAGA GCACCCATAC CTAACGAAAC 5100  
 10 GGCTACCCCA ACGATTGTCG CAAATACAGA TAAAACGTCA ATAAAAATCC CAATAGGACC 5160  
 TTCTACTTTA TCACCTAAAA GAGGACGTAA AGTTCTAGAT AATAAACCTG GTTCACCTTT 5220  
 ACGGAATTGC GAATATGCCA ACGTAACGCA ACAACACCAT AAACAGCCCA AGCATGGAAT 5280  
 15 CCCCAATGGA AAAATGTTGA ACGTAGAGCT TCAGTATAAG CTTCAGTAGT TTTGGGATCT 5340  
 GCTGTAGGTG GCGTAGCAAA GTGCGCCATC GGTTCAGCTG CACCATAAAA CACCAAACCT 5400  
 ATCCCCATAC CAGCACTAAA CAACATAGCA AACCATGAAA TTGTATTAAA CTCAGGTTTG 5460  
 20 TCATTTGGTT TACCTAGTTT AAGTTTTCCA ATAGGACTAA AAATAAGGAA TATACAGAAG 5520  
 AACACGATAA TCGTAGTAAG AATAAGATAA TACCAACCTA ACTTTTCTGT AATCCACATT 5580  
 25 TTAATATTAT TGTTAACATA GTTGAATTGT TCAGGTAAAA ATGCACCAAG TAATACGACT 5640  
 ATAGCAACAA CAATTGCACT ATAGATGAAG ACTGGTGAAT ACTTCTTTCC ATTGGATTTC 5700  
 TCTGGTGAAG AAGAATTCAT AATTAATTAC TCCCTTCAAT TCTATATTTA ATTTTATGTA 5760  
 30 GTAGAATAAA AATATTATCT AAACATTTTA TTCAATAACT CACG 5804

## (2) INFORMATION FOR SEQ ID NO: 250:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 35 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

ACCCGCGAAT ATGGTCCATC CTATCGATTT ATTTTAACT GGTTTGACAA TATTTAATTT 60  
 TTCATAATCA TTCTTAGTGA TTTTGACATA TGTTTTCGGT ATGAGCCAGT TAATAAATGG 120  
 45 AAAGAAGAAG ACAATCCAAT TACTTGCCAA ATCAATCATT AAATATTCAC TATCGTATTT 180  
 GATTATTCGA TATTTAGGGT TTTTATTAAT AACTTTAGAT TCGCAAAGCA ATGTCTCCAC 240  
 50 ATCCCTTTAA TTTTATGTGT AATACATTTT TCGATACTTC AAAAGACATT CAAATACTAT 300  
 CAAGTTACTG TCATCAAAGG TTTTATTAAC TGATATTTC ATATTTTAA TCTGAATTTA 360

55

## (2) INFORMATION FOR SEQ ID NO: 251:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 964 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

```

CCAGGGTGCG GAAAGCTTTA AAATTTGGAC TAATAAAGAT GCTGATATTA ATTCTATGAA      60
AACAGCAGTT TTACAACAAT TAAAAGGAGA ATAACATATG CTTACTGGCA AACAAAAAAG      120
ATACTTAAGA AGTTTAGCAC ACAATATTGA TCCGATTTTT CAAATTGGAA AAGGCGGTAT      180
CAACGAAAAT ATGATTAAAC AAATAGATGA TACGTTAGAA AACAGAGAAT TGATTAAAGT      240
ACATGTACTA CAAAATAACT TTGATGATAA AAAAGAATTA GCTGAAACAT TAAGCGAAGC      300
TACTCATAGT GAATTAGTGC AAGTGATTGG ATCTATGATA GTGATTTATA GAGAATCTAA      360
AGATAATAAA GAAATTGAAT TGCCATAATA ATGAAAAAGA TAACTTTTA CGGCGGTCAG      420
TTTAACCTTA TCCATACTGC ACATATGATA GTAGCTAGCG AAGTATTTCA TGAATTACAG      480
CCAGATGAAT TTTATTTTTT ACCTAGTTTT ATGTCTCCAT TGAAAAAGCA CCATGATTTT      540
ATAGACGTTT AGCACAGATT AACAAATGATA CAGATGATTA TCGACGAGCT TGGTTTTGGA      600
GATATTTGTG ACGATGAAAT TAAACGTGGT GGTCAAAGTT ATACCTATGA CACGATCAAG      660
GCATTCAAGG AGCAACACAA AGACAGTGAG TTGTACTTTG TTATTGGGAC GGATCAGTAT      720
AACCAACTAG AGAAATGGTA TCAAATTGAA TACTTAAAAG AAATGGTTAC TTTTGTAGTT      780
GTAAATCGAG ACAAAAATAG TCAAATGTT GAAAATGCTA TGATTGCAAT TCAGATACCT      840
AGGGTAGATA TAAGTTCGAC AATGATTCTG CAAAGAGTTA GTGAAGGGAA ATCTATCCAA      900
GTTCTTGTTT CTAAATCCGT TGAAACTAT ATTAAGGGGG AAGGATTATA TGAACATTGA      960
AAAA

```

## (2) INFORMATION FOR SEQ ID NO: 252:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1193 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

TTGTCATGAA ATAAATGGGA TGAATATCAC GACTAGAAGT AATGTTACGA ACAGGAGCGT 120  
 ATAAACTAGA GACGCTAAAT TCGACATAGT ATGThGCTCA ATTATGGCTG ATGATGAATT 180  
 5 TAAAGTATGT GCGTTGGAAC TGTCGGATTT TTGTTcATAA TGTTTTGcAT ATTGCGCCAT 240  
 GATGAATAGT GTAAAAATAA ATAAAACAAT AAGAGATATA ATGCCcATAA TCAAAAGTAT 300  
 TTGTTTAGAG CCTTTCATTA TTTCACATCC TTTCTAAAAT ATATTTGTAA CTAAATTTAA 360  
 10 AATAGTTATT TTTGTAATTC TAAACCTTTT TCATCGCGAA AACAATTAAA TAGGTcGCGG 420  
 TATTAATTAT TATATTATTA CCGCTTAATA TGAAAAATAC ATGAAaAATTA ATTTTCTAAT 480  
 ATACTTTTGA AAAATTATTA CAAATTAGCC CCTTCAAAAC GCGAAAACAT AAGGATTCTA 540  
 15 GTTTCAAAAG GGCTGATAAG CATAAAATGA AATGTAATAT TCGATGTAT AAAATTTTTA 600  
 ATTAGCTAAA AATCATCGCA TTAATTTTTT GAGCTACATC ATCAAAATTC GGACATTTTA 660  
 ACGACACATA TAATTTAATT TTAGGTTcAG TACCAGAAGG ACGTAAAGCG ATAAATCCTT 720  
 20 CGTCAAATAA GACACGAATA ACATTTGATT TAGGAGAATT AATCTGCGAC GTTGTATCTT 780  
 TATCCAAATG ATAAACCTCG CTAGTTAAAT AATCTTCAAT TGCTTCACT TTGAGTCCTT 840  
 25 GAATCTCTTG CCGTGGATTT GAACGGAATT TGGTCATTAT TGCATTAAAT TTCTTTTTCC 900  
 CTTCAAATCC TTCTAGCGTA TGCGgAATAA TGTATCCTCA TGTCTACCAa CAGTTTGATA 960  
 AATCTGTTCT AATTCATCTT TCAATGTTTT GCCATATAAT TTAACTCAG AAGCGTATTT 1020  
 30 TATAATGAGT GGCACAATTT GTACGGCATC TTTATCACGT ACAAAGGCT CTGATAGAAA 1080  
 ACCGTAActC TCTTCAAATG CGAAAATCAT ATTTGATGAT CATCCAGTTG TCTTATTTCC 1140  
 TGAGCAATAA ATTTAAGCCC GTCAGCACCT CTTTGGTATT CAACATTATT ATA 1193

(2) INFORMATION FOR SEQ ID NO: 253:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1098 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

TATCAGGATG ATTATGhCGG GGTTTTATTA AGTCTGAGCG TATAThCCTT TTATTCTCCT 60  
 AAATGGTTTG GAATTCAAGG AGAAGACGAT GAAATGGTAT CCAAAAACCA AnGAATATAT 120  
 50 hGCATTAAAGT CTGAGGATGA TAGTGCGGTG GCAATTCGTT CATTAATTTT GCATAAAGAT 180  
 GAACCTATGT ATTTAAAAAA ACGTACATGT GTACCTACTT TGTTAATTAA TGGGGAACAT 240

AAAAAAATCT TCGAACATTC AGGACATGCA CCGCATATTG AAGAACCAGA AGCATTATG 360  
 AATTATTATT TAAAATTTT AAAAAGCGTA TCATAATATG TGATATATAA ACCTAGGGCA 420  
 5 TAAAGTCCTT AGGCAATGTG AAAAAGCTGA TTACTATTCA TTATTTGATA GAAATCAGCT 480  
 TTTTGTGAAA TGTATTTGAT ATATACTGCT CGTTATGCGG CTATCTTCCT TATATTAAGT 540  
 GCCATTAGTG CAAAACCTCT TAACAATTAG GTAAAAAGAG CATAAAAAA GGAAGTTTAA 600  
 10 TAGAATGTAT CATCTATCAA ACTTCACCAA ATTGCGCTAA ACAAATTAT AGTTCAATTT 660  
 CGTTGTTTGC TTCAGTGATT CGTTTATTTA CTCGACTCAA TAATGATTCTG ATTTTTTTAC 720  
 GTTGTGTGTC ATTAACAAGA ATTAATACAG TTCTTTCATC ATGCTCATTG CGTTTTTTAT 780  
 15 CGAAGTAATC TTCTTGAGAT AAAATTTTAA CTGCTTTAAC AACTTGTTGGT TGTTTGAGT 840  
 TTAAATGATT AATAATATCT TTAAGATAGT ATTCTTTCTC TTTGTTTTCTG CTGATGTATG 900  
 20 TCAATACAGC GAATTCTTCA AAGCTAATTG AATATTCCTT TTTAATTAAA CTTTTTAATT 960  
 TGTCAGCATA AGTGACCATT GATAACAACT CAAAGCAATC ATTGATTTTT GTAATTGCCA 1020  
 TGTTTAAAC CTCCCTATTT GATGCATCTT GCTCGATACA TTTGCCCCGA TAATATATG 1080  
 25 TATCTAATCT TTATGnAT 1098

(2) INFORMATION FOR SEQ ID NO: 254:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2881 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

CCAAGTAAAA TTGTGCAATC ATTTGACGCA TTAATGGACG CCTTGGACAA TGAAGATTAT 60  
 GAAGGAGAAA AAGTCATTCC ATTCTTAGAT AAACATTTTA AATATCAAGA TGGCCGATCA 120  
 40 AGTGAGCGTT TAGTCAGAAA TTTATTTGGT AGCTAAGTTT ATATAGTAGT CAAAGTGGA 180  
 GAGGTATAAT GATGAAATTT TCAGTAATAG TTCCAACATr CAATTCAGAA AAGTATATAA 240  
 CAGAATTACT TAATAGCCTT GCGAAACAAG ATTTTCCGAA AACTGAATTT GAAGTGGTTG 300  
 45 TAGTTGATGA CTGTTCAACA GATCAAACGT TACAAATAGT TGAAAAGTAT CGCAATAAAT 360  
 TGAAGTTGAA AGTAAGTCAA CTCGAAACAA ATTCTGGTGG TCCAGGTAAA CCTAGAAATG 420  
 50 KGGCGTTAAA ACAAGCAGAA GGTGAATTTG TATTATTTGT GGAATCCGAT GACTATATAA 480  
 ACAAGAGAC TTAAAGGAT GCAGCAGCAT TTATTGATGA ACATCACTCA GATGTCTTAT 540

## EP 0 786 519 A2

	CACCTGAAGT TACTTTGTTA AATTCAAGAA TTATCTATAC TTTAAGCCCG ACTAAAATCT	660
	ATAGAACAGC ATTACTAAAA GATAATGACA TTTATTTTCC AGAAGAATTA AAGAGTGCAG	720
5	AAGATCAATT ATTTACAATG AAAGCATATT TAAATGCAAA TCGAATCAGT GTGTTAAGTG	780
	ATAAAGCGTA TTATTATGCT ACAAAGCGTG AAGGTGAACA TATGAGTAGT GCGTATGTTT	840
	CACCTGAAGA CTTTTATGAA GTCATGAGAT TGATTGCTGT AGAAATATTA AATGCAGATT	900
10	TAGAAGAAGC CCATAAAAAT CAAATCTTAG CAGAATTTTT AAATCGTCAT TTTAGTTTTT	960
	CTCGTACGAA TGGCTTCTCA CTTAAAGTTA AACTAGAAGA TCAACCACAA TGGATTAATG	1020
	CTCTAGGAGA CTTTATACAA GCAGTTCAG AACGTGTAGA TGCATTGGTG ATGAGTAAAT	1080
15	TACGACCATT GTTGCACTAC GCGAGAGCGA AAGATATAGA CAACTATAGA ACTGTGGAAG	1140
	AAAGTTACCG TCAAGGTCAA TACTACCGTT TTGATATTGT AGATGGTAAA TTAAACATTC	1200
20	AATTCAATGA AGGCGAACCA TACTTTAAAG GCATTGATAT CGCTAAGCCA AAAGTGAAAA	1260
	TGACAGCATT TAAATTTGAT AATCATAAAA TTGTTACAGA GCTAACGTTA AATGAATTTA	1320
	TGATTGGCGA AGGACATTAT GATGTCAGAC TTAAATTACA TTCACGAAAC AAGAAGCACA	1380
25	CAATGTATGT ACCTTTAAGT GTCAATGCGA ATAAACAATA TCGTTTTAAC ATTATGTTAG	1440
	AAGATATTAA AGCGTATTTA CCTAAAGAAA AAATTTGGGA TGTTTTCTTA GAAGTCCAAA	1500
	TAGGTACGGA AGTATTTGAA GTGCGTGTG GTAATCAACG TAATAAATAT GCATATACTG	1560
30	CAGAAACAAG TGCATTAATT CATTTGAATA ATGATTTTTA TAGATTAACA CCGTATTTCA	1620
	CAAAAGACTT TAATAACATT TCGTTATACT TTACAGCTAT TACATTAACG GATTCAATCT	1680
	CATTGAAGTT AAAAGGTAAA AACAAAATCA TTTAACTGG TCTGGATCGT GGTATGTAT	1740
35	TTGAAGAAGG TATGGCTAGT GTCGTACTAA AAGACGACAT GGTGATGGGA ATGTTAAGCC	1800
	AAACATCAGA AAACGAAGTG CAAAATCTTA CTTAGCAAAG ATATTAAAAA GCGAGACTTC	1860
	AAAAATATTG TTAAGTTAAA CACTGCACAT ATCACTTATC CACTAAATAA ATAATAAATG	1920
40	CCCTCAAATC ATTGTGAGCC AACATGATTT GAGGGCTTTA TTTTGCTGTT TATGACATGA	1980
	TTATGACATT TCCCTGATTT TCATTTTCAT ATACATTAAA TTGTATACAC TGGAAATGAG	2040
	GAGGTTATCT ATAATGATAA ATAAAAATGA CATAGTAGCA GATGTAGTAA CTGATTATCC	2100
45	GAAAGCAGCG GATATTTTTA GAAGTGTGGG AATAGATTTT TGTTGTGGCG GACAAGTAAG	2160
	TATAGAAGCA GCAGCCTTAG AAAAGAAAAA TGTAATTTG AACGAATTAT TACAGCGTCT	2220
50	CAACGACGTT GAACAAACGA ATACACCAGG TTCGTTAAAT CCTAAATTTT TAAATGTTTC	2280
	ATCACTTATT CAATATATTC AATCAGCATA TCATGAACCT CTAAGAGAAG AATTTAAAAA	2340

55

TGAGTTAAAA GAAACATACG ATACATTTAA AAATGGCATG TTAGAGCATA TGCAAAAAGA 2460  
 AGACGATGTC GATTTTCCAA AACTCATTAA ATATGAGCAA GGTGAGGTAG TAGACGATAT 2520  
 5 TAATACTGTG ATAGATGATT TAGTTTCAGA CCACATTGCA ACGGGAGAAT TGTTAGTAAA 2580  
 AATGAGCGAA TTAACATCTA GTTATGAACC TCCGATAGAA GCGTGTGGTA CTTGGCGACT 2640  
 TGTITATCAG AGATTAAAAG CACTTGAAGT GTTAACACAT GAACACGTAC ATTTAGAGAA 2700  
 10 TCACGTATTA TTTAAAAAAG TATCATAAAT AACGCGATTA GAAACTGTTG GCAAAAATAA 2760  
 GTCCAGCAGT TTTTCGCTAT GTATAAAAGT CATAATAGTG ACATAAACAG CATTATTTGA 2820  
 AAAGAAaAAT GGTCAACTTA GCATAAAAAT TGATATGAAn ATTTAATGGT ATAGATAATT 2880  
 15 A 2881

## (2) INFORMATION FOR SEQ ID NO: 255:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1056 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

ACCGTCGAAT ATCGCTTGTTG ATTTACAATT TGTGTATTAA GATGCTCAAC TAATTGCGGT 60  
 30 ACATATTCCG AATTAGATT TGCAAGTACA ACAATTCCAT AATTTGTTT TGGATTAGT 120  
 AAAATAAATG ATGAAAAGTT ATCTAGCGTT CCTGAATGAA AACTAAATG TTCATCATT 180  
 TTGGTAAACC AGCCGAAGC ATATGCATTG GCATTAGGTT CACCAATTGT TGAAGATAAA 240  
 35 TTTTATGTG ATTGTTGAAC TAATGATTTG TATTTATCAG GTGGATTAAG TTGAATTTT 300  
 ATCGAATGTT CCAAATCTTC AGTTGATGTC ATCATATATG CTGATGGTGT ATCCCAAAGG 360  
 TTAAATTCAG GTTTAGAGAC GACAGGTGTC GAACCTTGTA ATTCATAGCC AATAGCATCA 420  
 40 TGTTTTGATT TGTAATTGGT TTGTTTGAAT GATGTATGTG TCATATGCAA AGGCTTGAGC 480  
 CATGAATTTG TAATATATTT TGTATAGGAT TGCTTCGTAA CGTTTTGGAT AATTAAACCT 540  
 AATAATCAT AGTTCATATT TGAGTATTCA AATTCTTCTC CGGGCTTATG ATGTAATTCA 600  
 45 TCACCCATAA TTGCATGGGT TACATCATTT AAACGATTAT TTTTGCTTGT CACAGAATCT 660  
 TCGCTTGTA TATCACTAGG TATACCACTT GTTTGAGCCA AAAGTTGCTT AATCGTAATA 720  
 50 GTTTCATTTT GACCATTATA GTTCATTTTA AAATGAGGCA CATGTTTGA TACGGCATCA 780  
 TTTAAGTTTA ATCGACCTTC TTGAGCTAAT TTTAAATG CAAGACCTGT GAAAGCTTTC 840

55



TGATAACCAT AACCTTTATT TAAAAAACT TTGCCATTTT TTACTAyTAA AATTGATGCT 960  
 CCAGGAATGT GTCCCTTTTG TAAATCATGC TCGATAATTG TATCTATTG TTGTTGCGAA 1020  
 5 TCATTGGTTA ACCGTGTCTT CGTATTGCTA TTTAAT 1056

(2) INFORMATION FOR SEQ ID NO: 256:

(i) SEQUENCE CHARACTERISTICS:

10 (A) LENGTH: 1277 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

ATGCCaCACT TATTGGTGGC AGGATCGACG GGTAGTGGTA AATCTGTTTG TATAAATGGT 60  
 ATTATTACAA GTATTTTATT AAATGCTAAG CCGCATGAAG TTAAACTTAT GTTAATCGAT 120  
 20 CCGAAAATGG TTGAACTAAA TGTTTATAAC GGAATCCAC ACTTATTAAT TCCGGTTGTT 180  
 ACAAATCCTC ATAAAGCTGC TCAAGCTTTA GAAAAAATTG TAGCTGAGAT GGAAAGACGT 240  
 25 TATGATTTAT TCCAACATTC ATCAACTAGA AACATTAAAG GTTATAACGA ATTAATCCGT 300  
 AAGCAAAATC AAGAATTAGA TGAGAAGCAA CCAGAATTAC CTTATATCGT TGTTATTGTA 360  
 GATGAGCTTG CAGATTTAAT GATGGTAGCT GGTAAAGAAG TTGAAAATGC GATTCAACGT 420  
 30 ATTACACAAA TGGCACGTGC AGCAGGTATA CATTTAATTG TAGCGACACA AAGACCTTCT 480  
 GTGGATGTAA TTACAGGTAT CATTAAAAAT AATATTCCAT CTAGAATAGC TTTTGCTGTG 540  
 AGTTCTCAAA CAGATTCAAG AACTATTATT GGTACTGGCG GCGCAGAAAA GtKACTTGGT 600  
 35 AAAGGTGACA TGTTATACGT TGGAATGGT GACTCATCAC AAACACGTAT TCAAGGGGCG 660  
 TTTTAAAGTG ACCAAGAGGT GCAAGATGTT GTAAATTATG TAGTAGAACA ACAACAGGCA 720  
 AATTATGTAA AAGAAATGGA ACCAGATGCA CCAGTGGATA AATCGGAAAT GAAAAGTGAA 780  
 40 GATGCTTTAT ATGaTGAAGC GTATTTGTTT GTTGTGgAC AACAAAAGGC aAGTACATCA 840  
 TTGTTACAAC GCCAATTTaG AATTGGtTAT AATAGAGCAT CTAGGTTGAT GGATGATTTA 900  
 GAACGCAATC AGGTAATCGG TCCACAAAAA GGAAGCAAGC CTAGACAAGT TTTAATAGAT 960  
 45 CTTAATAATG ACGAGGTGTA AAAAAATGTC AGAAATGAAT GCGGTATATA ACGTTAAACA 1020  
 ATaCATTtTa AATTTgATTA AGCAAAATAA ATTGGAATAT GGTGACCAAC TTCCAAGTAA 1080  
 50 TTTATCAATT GCCAGAGAAT TAAATGTAAA AACCGACGAT GTTTATGAAG CAATTCAGcA 1140  
 TTGATTACTG AACAAAGTCAT TAAAGATaT TTGAAGAGGG CACAAGTgTT AAGTCACTGC 1200

55

GrTTGAATG CGGAAC

1277

(2) INFORMATION FOR SEQ ID NO: 257:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3557 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

15 TACCGTCGTT TTATGyGTCA AATTTTACAG TAAATTTTGC TTCATCAAAA GAAATAACCT 60  
 TTAACAAGTA TAGTAATTTT ACATTTTACA ATGTTACAAA ATATAATTTT TTATAATTAG 120  
 TTAATATCAC TAAAACGCTT TTATACACTA TCAATCAGC ATTTATAAAA ATATGAACCG 180  
 20 ATATCTATAA ATGTTAATAA TATTACAAGA TAATAACAA CCACACAAAG CTACTTATTT 240  
 TTGATAATAT GGAAATCGTA ATATAAAACn AAACTTAAT TTACTATATA AATTGTCTTA 300  
 ATAATTTTAA AAAGTAGTAA AACATAATTT TAAGGAGGAG TCCCTTTGAA AAAATTAGCA 360  
 25 TTTGCAATAA CAGCAACATC TGGTGAGCT GCATTTTAA CGCATCATGA TGCACAAGCT 420  
 TCTACACAAC ATACAGTACA ATCTGGTGAA TCATTATGGA GTATTGCTCA AAAATACAAC 480  
 ACTTCAGTAG AGAGTATTAA ACAAATAAAC CAATTAGATA ACAACTTGGT ATTCCTGGT 540  
 30 CAAGTTATCT CAGTAGGTGG AAGTGATGCA CAAATACGT CAAACACTTC TCCACAAGCT 600  
 GGTTCAGCAT CATCTCATAC TGTACAAGCT GGTGAATCAT TAAATATCAT TGCTAGCAGA 660  
 TATGGTGTTC CAGTTGATCA ATTAATGGCA GCCAATAACT TACGTGGTTA TTTAATTATG 720  
 35 CCTAACCAAA CATTACAAAT TCCTAATGGT GGATCAGGTG GTACAACACC AACAGCTACA 780  
 ACAGCTAGCA ATGGCAATGC ATCATCTTTT AATCACCAAA ATTTATACAC TGCTGGTCAA 840  
 TGTACATGGT ACGTATTTGA CCGTCGTGCT CAAGCTGGTA GTCCAATTAG CACATATTGG 900  
 40 TCAGACGCTA AGTATTGGGC TGGTAACGCA GCTAATGATG GTTACCAAGT AAACAACACA 960  
 CCATCAGTTG GTTCAATTAT GCAAAGCACA CCTGGTCCAT ATGGTCATGT TGCTTATGTT 1020  
 GAACGTGTCA ATGGTGATGG TAGTATCTTG ATTTCTGAAA TGAATTACAC ATATGGTCCA 1080  
 45 TACAATATGA ACTACCGTAC AATTCCAGCT TCAGAAGTTT CTAGCTATGC ATTCATCCAT 1140  
 TAATTAAATA AATTGTACTG ATATATACTA GCAATTCACA TCATGTGAGA TTGCTAGTTT 1200  
 50 TTTATTTTGG AAAAAAATTT TCATTTTGGT AAAAAAATT ATCTCACCCT TCCCTATCAT 1260  
 ACATATTTAT ATTTGTATG AATGGTAGTT AGGTAAAAAT TAACAACCTA CCTATTTGAT 1320

55

	ATTTAATTG TTATACCAGT ATTTTACGCT TTTTCGTCTA CATATACAAA TTTATATTAA	1440
	ATAAAGCCCA ATACAATTTA GGTTAATTAA ACAAGTTGAT AACTATTTAA TTATTCCTTC	1500
5	ATTGAAGAAT ATAAACTATT AAATCATTAT TTTGCTCTTA CATATATTTT AATGACCTAA	1560
	CTGaTTATGT TCCATGGAAT ACATTTATAA TATAGCCTCC TAATTAATAT GCyTTGTCTT	1620
	GGTCATTCTA CGTAAATTCT ATAAAAATAG TTATCTACTT ACATAAATAT CTGrACTTCA	1680
10	ATACCACCAT ATGTTTGTGA TACTGAAGTT CAGTTTAGTT TTATTTTCAA TTAGAAAAAT	1740
	AAGTTAAGTA TATAGAATAG TAAACCTGCT AACAAATGCTG AAATAGGTAA TGTAAATCACC	1800
	CATGTAATGA TCATTCGTTG CGCAGTGCTC CATTTTACAC CTTTAGCTCG GTTAGAAGCA	1860
15	CCAACACCTA AGATTGATGA TGACACAACG TGAGTTGTTG ATAATGGGAA ATGTAGCGAT	1920
	GATGCAACAA AAATTGTAA TGCAGATGAT AAATCGGCCG CAGCACCATT TGCTGGACGT	1980
20	ATTTTCATAA TATTACCACC TACAGTTTGG ATAATTTTCC AGCCACCAAT TGCAGTACCA	2040
	AGCCCCATTG CTGTGCGACA GGCAAATTTT ACCCATAACT GTGGTTCAAC ACTGCCATCA	2100
	TTCTGTACAT TAGCGACAAT CAATGCCAAC GTAATAATAC CCATTGATTT TTGCGCATCA	2160
25	TTCGTACCGT GAGAGAATGA TTGTAACGCT GCTGTGAAAA TTTGGAAAAA TCTAAAGTTA	2220
	CGATTCGCTC TTGTTAAATT TGCATTTTTA AAGATAACTT TAAAAATTGA ATACATCAAG	2280
	AAACCAACAC AAAATGCGAT AATCGGTGAA ACGATTAATA CAATAATAAT TTTTGTGAAA	2340
30	CCTTGGTAAT GTAACACTCC AAATGAGCCT TCAGATGCGA TTGCTGCACC CGCAATTGAA	2400
	CCTATAAGTG CATGTGAAGA CGAACTTGA APTCCGTAAA ACCAAGTAGC TAAATTCCAA	2460
	ATAATAGCCG CAAGTATTGC AGCTAACACA ACAACTAATC CATTTTCCAA TTAAATGGA	2520
35	TCGACAATGT CTTTAGTAAT GGTGCCTGCA ACGCCCGTAA ATGTTAAAGC ACCTATAAAG	2580
	TTCACTCACTG CTGCCATTAA AATTGCCGTT TTAGGgTTAA CGCTCTAGTA GATACAGCAG	2640
	TAGCTACTGC ATTGGcTGTA TCATGGAaTC CcATTGATAA AGTCAAATAT CAGCGAGAAA	2700
40	ATAACTACAG CTATAGTGAC GATGATTATA TATGACATAA ATATATACTC CCCTTAGCTA	2760
	TTTTTCATAA TAATAGTTTC AAAATTATTT GCTACGATTT GACATTTATC AGCGATTTCT	2820
	TCCATGCTTT CATAAATATC TTTTATTTTA ATTAAAGTGA TTGGATCTGT TTCGCTATTG	2880
45	AAAATATGTT TAATTGACTG TCTTAAAATA CCATCACAGT TTGTTTCAAA TTCTTTAATA	2940
	TTAATTGAAT GAATACGCAT ATGTGATAAT TTTTATCGA CTAATAAGCC GACAGCAAGT	3000
50	TTCATTTCTG CAACTGCTTT TTGAATGTTA TCAACAAACT CAGCCATATA TTCATCTGTG	3060
	TATTCGATTG AATACATTTT AAACATrGCT GCCGTTTCTT CAATTGCATC TAAAACATCA	3120

55

TTTAAATCAG TAATTACTTG ATGTACTAAt tCGcACCATG TGA CTCATAA GTTTTAATGT 3240  
 TGTCTGAGTA TGCTTTTAAA TCTAAATGTG TATTGAAATC CATTTTACCG AATTCAATAG 3300  
 5 CAGCAGGATC CAGATTGAAA ACCATCTCTT CTAATTGAAC CATAAACTTA TCTTTTCT 3360  
 TACTAAACAT TTAAAATCCT CCATTTAAGC GATTGTCACC AATCACATTC AGTTATAATT 3420  
 TGTTCAAAT TAAGACAAGT GAATTTACAA ACTAATGATA CAAATTTGTT ATTATCAATC 3480  
 10 GTCAGTATAA TTTTAGTGTA CTGATATTAA TTTCAAAAAT GCCTCACAGT AAACAATTTA 3540  
 CTGTATTTGC CCTTATA 3557

## (2) INFORMATION FOR SEQ ID NO: 258:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1631 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

25 AACTATACAT TTCGAAAAAT TCTTCTAGTG AACCTGCGCC ACCAGGAGCC ATGACAAATG 60  
 CATCTGCAAG TTCTGCCATT TTATTTTnAC GTTCATGCAT AGAATCAACT AAAATTAATT 120  
 CAGTTAAACG TTGGCTTG TG ATTTCATGTT CATCTAACAT TTTAGGCATG ACGCCAATAG 180  
 30 CTTTGCCGCC ATGATCTAAT ACACCATCTT GaATGGCACC CATAATGCCA ATTGACCCTG 240  
 CACCAAATAC TAATTCATAA CCTTGTTT CAG CAAAATATTT ACCTAAATCG TATGCTTTTT 300  
 GTACATATGA AGGGTCATGA CCTTTGCTTG CACCACAATA AACTGCGATT CGTTTCATGT 360  
 35 TAATCCAGCT CCTTAATTCG ATGAATGACT TTTAATAGTG ATTGTTCAAA CACTTTTGA 420  
 TCTTGCTTTG TAAAAGGTGG GGGACCTTTG TGGCGACCAC CTTGTTTTCT AATTTGTGCA 480  
 TTCATATATC GTTTATCTAA TAGTTGTGTA ATATTTTGG AATTGTATAT CTTCCCATTA 540  
 40 TGATGCATGA CAATTAAGAC TTTGTCGACT AATAAACTTG CGAGTCCATA ATCTTGAGTG 600  
 ACTACGATAT CATCCTTCGT TGATAATTGA ACAATTTTGT AATCAACTGC ATCTGGTCCA 660  
 TCATCAACAT ATAATGTTGA TACATGTGGA GGATATAATT GGTTCGAAAA ATGGCTGAAG 720  
 45 CTCCGAATAA TTGTCACAAA AATGCCTGTC TCAGTTGTTA AATCTATAAT AGAATCAACA 780  
 ACAGGACAAG CATCTCCATC AATAATAATA TGTGTCACAA TTATGCCTCT GTATTGTTTT 840  
 50 CTTTATTTTG TTGAGAGGCG CTTTGGCAA CATAATCTTT ATATTTTTTA AATGACTTGA 900  
 TGCCTGCTTT ATCAGCTTCT TGTTGGCGTT TTTGTTCTTC TTTGTGTCGT TTTTCAATAT 960

55

CGCCTTTTTT CTCAGTTTTT TCATCTAATT TATTAGGTGT TAAGCCTGCT TTTTCTTCGT 1080  
 ATTTTGTGTA TTTTTCATA TCTTTAATAC GTTGTATTTT ATTCTTTTCG CGGGCTTTTT 1140  
 5 GCTCTTCTTT ATGACGCTTT TCGATATTTT TTTGAAGTAT TTTATTCATT TTATCAGCGT 1200  
 CTTTACGATT TTGTTTAGCT AATTTTTTCG CTTTTTCTC AATATAGGCA GGATCATGTT 1260  
 CTCTAGCAAA CTTTTTAAGT TCACGTTTAT TTTCAAAATC TTGTTTTTTA TCGCCGACAT 1320  
 10 ATTCTTTAAC ATCACTCGCT GTGTTACTGA TTGCTGCAGA TGTTTTTGAA GCAACTTTAC 1380  
 TTGTAGCATC TGTAACTTTT TGTACGTCGG GATGTTGTTT GATACGTTTA CGTTCAACAA 1440  
 TTAACGGTAC CAATACAATT GGTAATACAT TAATCATAAA TTTGATGACT TTTTCTTAT 1500  
 15 CCATAGATCT TGcCTCCaTA ATTACTTTAT TAAtTTTACa TACCcTATGa TACATCAATA 1560  
 TAAACGATGA TAGTAGTGAA TCACTATTAA GTATTTcAGA TGTTTtTTAA AAgAAGaCCC 1620  
 AATTAnAAAA A 1631  
 20

## (2) INFORMATION FOR SEQ ID NO: 259:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6645 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

CGAAATCATG ATTTAATGCT TTTTCATATA AGCTTTTCCA ATTAATCTTT CGTCCATGAT 60  
 ATTCTTCAAC TGTTGCTAGA TATTGTGCAA TTTTAGTTAC TTAAAGGAG TGTGCTGCAA 120  
 35 CaTTGTGkTC mAAATATTTA AATTTTCCaG GtAAATCTTAT AAGTCTTTCC aTATCTGATA 180  
 ATCTtTTAAA ATATTGATGT ACACCCATTT CAATTACCTC CTCCATTAAAT TAATCATAAA 240  
 TTATACTTTC TTTTACATA TCAATCAATT AAATATCATT TaaATATCTT CTTTaTATAA 300  
 40 cTCTGATTAA ATGATACCAA AAAATCctCT CAACCTGTTA CTAAACAGG CTAAGAGGAT 360  
 AGTCTTGTCT TGATATATTA CTTAGTGGAT GTAATTATAT TTTCTGGAT TTAAAATTGT 420  
 TCTTGAAGAT TTAACATTAA ATCCAGCATA GTTCATTTCA GAAACAGTAA TTGTTCCATT 480  
 45 AGGGTTTACA GATTCAACAA CACCAACATG TCCATATGGA CCAGCAGCTG TTTGGAAAAT 540  
 AGCGCCAAC TCTGGTGTtT TATCTACTTT AAATCCTGCA ACTTTTGCTG CGTAATTCCA 600  
 50 GTTATTGCA TTGCCCCATA AACTTCCTAT ACTTCTACCT AATTGTGCAC GACGATCGAA 660  
 AGCATAATAT GTGCAGTTTC CATAAGCATA TAAGTTTCCT CTGTTAGCAA CTGATTTATT 720

55

	TACATTA	ACT	GTCTTAG	TTA	CTGCTTG	CCTT	AGGTG	CCTGC	TTAACTA	CTA	CTTTTT	TAGA	840
	TGCTTG	TGT	ACAGGT	TGT	TTACTA	CCTT	TTTAG	CCTTG	CTTGCT	TTTC	TTACTG	GTGA	900
5	TTTAAC	CGCT	TTAGT	TTGT	TCACT	TTATT	TTGAG	GCACA	AGTGAA	TCA	CGTACC	CAGG	960
	AAAAAT	TAAA	GGTG	TACAC	CAGGAT	TGTA	TGAAT	TATAAT	TGATT	CAACG	TTAAG	TGATG	1020
	CTCTAA	AGCA	ATCTT	AATA	ATGAAT	CGCC	AGCAAC	TACT	GTATA	AGTTG	TCGGT	GATTG	1080
10	CGTTTG	TGCT	TGAAC	ATTG	ATACATA	AATT	ATGTT	GAAACA	GGTGT	TTTTTA	CTTGT	TGCC	1140
	ATGTTG	TGT	GCATG	TGCTG	CATTAT	TTAA	AGCTA	AAAAA	GCTAAC	ACTG	ACGAA	ACCGT	1200
	CACTGT	AAGA	GATTTTT	TCA	TCTTG	CTGTC	ATTCCT	TTGC	TGTTAG	TATT	TTAAG	TATGC	1260
15	AAATACT	AATA	GCACA	AATA	TTTTG	TCAA	AGCTAT	TGTT	ATAAC	GATGT	AATCAA	TATGG	1320
	TTAACA	AATAT	AAAAA	GAATA	CAACCT	TTTA	TCATAG	TGTA	AAATG	TATTC	ATACCA	TGTA	1380
20	ATTGAGA	ACG	TTTTCA	AATA	TTAATT	CAAT	ACCTT	GAAAA	TCGCC	CATAGG	TAATAT	TACT	1440
	AAATGC	CACAC	TGCAT	ATGTT	GTTTTA	ACAA	ACACA	ACTTT	TAAAAA	AATAT	ATTCTA	ACTC	1500
	TATCTA	CCGA	ATTGT	ACTTA	AATAT	TATA	AACAA	TATCAT	ATTCCA	AAAT	CTAAT	TACA	1560
25	ATTTAT	TTTAG	CTACCT	TTTA	AAAAAC	CAAA	AACCG	ACGCC	CTTTT	AGAGC	CTCGG	TTTA	1620
	AAATAT	AATAT	TAATC	GTGCG	ACATT	GTCTG	TCTTAA	AATAT	GATT	CGATAA	ATGGT	CCAAT	1680
	GTCTCC	CATCC	ATCACT	GCAT	CAACCT	TACC	TGTTT	CCTCG	TTCGT	ACGAT	GATCT	TTCAC	1740
30	CATTGAG	TAT	GGATG	GAAAA	CATAT	GATCT	AATTT	GGCTT	CCCCA	GCCGA	TTTCT	TTTTG	1800
	TTCCGC	CACGA	ATTT	CAGCCA	TTTCA	CGTGC	CTGCT	CCTCC	AATTTT	AATT	GATATA	AATTT	1860
	AGACTT	TAAAC	ATTTT	CATAG	CTGCT	TACG	GTTTT	TAAAT	TGAGA	ACGTT	CATTT	TGGTT	1920
35	ATTAACA	ACT	ATACCT	GAGG	GGTGG	TGGGT	AATTC	GTATT	GCCG	ATTCAG	TTTTG	TAAAT	1980
	ATGCTG	ACCA	CCTGC	ACCAG	AAGCT	CTGAA	TGTAT	CAACT	GTAAT	ATCAT	CCGG	ATTGAT	2040
	TTCAAT	CTCT	ATTT	CATCAT	TATTA	AAATC	TGGA	ATAACG	TCGC	ATGATG	CAAAT	GATGT	2100
40	ATGACG	ACGT	CCTG	ATGAAT	CAAAT	GGAGA	AATTC	GTAAT	AGTC	GGTGTA	CACCT	TTTTC	2160
	AGCTTT	TAAA	TAACCA	TAAAG	CATTAT	GCCC	TTTGAT	GAGC	AATGT	TACAC	TTTTA	ATCCC	2220
45	CGCTTC	CATCC	CCAGG	TAGAT	AATCA	ACAGT	TTCAAC	TTTA	AAGCCT	TTTCT	TCTCA	CAATA	2280
	ACGTTG	AATAC	ATTCT	AAATA	GCATAT	TAGC	CCAAT	CCTTGA	GACTC	CGTGC	CACCT	GCACC	2340
	AGGATG	TAAC	TCTAG	AATTG	CGTTAT	TGGC	ATCGT	GAGGC	CCATC	TAATA	ATAAT	TGCAA	2400
50	TTCGT	ATTCA	TCCA	TTTAG	CCTTAA	AAT	AATG	ACCTCT	TGCTC	TAAGT	CTTCT	TTTCAT	2460
	TTCTTC	CATCA	AATTC	TCTCT	GTAAT	AAATC	CCAAG	TAGCA	TCCAT	GTTCAT	CTACT	TCTGC	2520

55

	TTGCGCTTTC GTTTGGTTAT CCCAAAAATT AGGTTCTGCC ATCATTCTT CATATTCTTG	2640
	AATATTAGTT TCTTTGTTCT CTAAGTCAAA GAGACCCCT AATTGTGTT AAATCTTGAT	2700
5	TATACTTATC TATATTTCGT TTGATTTCTG ATAATTCCAT AGCATTGCT CCTATTTATA	2760
	TTTCAATTCA AGTCATTGAT TTGCATCTTT TATAATGCTA AATTTAACA TAATTTTGT	2820
	AAATAACAAT GTTAAGAAAT ATAAGCACAC TGACAATTAG TTTATGCATT TATTGTTAAA	2880
10	AATCAGTACA TTTATCATCG ACATATGCCT AAACCGATT TTTAAACTA AGTACATAAC	2940
	AACGTTTAAAC AACTTCTTCA CATTTTTTAA AGTATTTAAC GCTTGTAATA TAAAAAGACT	3000
15	CCTCCCATAA CACAACTAT AGGTGTTTAA TTGGAAGGAG TTATTTTATA TCATTTATTT	3060
	TCCATGGCAA TTTTGAATT TTTTACCACT ACCACATGGA CAATCATCGT TACGACCAAC	3120
	TTGATCGCCT TTAACGATTG GTTTCGGTTT CACTTTTTCT TTACCATCTT CAGCTGAAAC	3180
20	GTGCTTCGCT TCACCAACT CTGTTGTTTT TTCACGTTCA ATATTATCTT CAACTTGATC	3240
	TACAGATTTT AAAATGAATT TACAAGTATC TTCTTCAATA TTTTGCATCA TGATATCAAA	3300
	TAATTCATGA CCTTCATTTT GATAGTCACG TAATGGATTT TGTTGTGCAT AAGAACGTAA	3360
25	GTGAATACCT TGACGTAATT GATCCATTGT GTCGATATGA TCAGTCCAAT GGCTATCAAT	3420
	AGAACGAAGT AAAATCATAC GCTCAAACCTC ATTCATTTGT TCTTCTAAGA TATCTTTTTG	3480
	ACTTTGATAT GCTGCTTCAA TCTTAGCCCA AACGACTTCG AAAATATCTT CAGCATCTTT	3540
30	ACCTTTGATA TCATCCTCTG TAATGTCACC TTCTTGTAAG AAGATGTCAT TAATGTAGTC	3600
	GATGAATGGT TGATATTCAG GCTCGTCATC TGCTGTATTA ATATAGTAAT TGATACTACG	3660
	TTGTAAAGTT GAACGTAGCA TTGCATCTAC AACTTGAGAG CTGTCTTCTT CATCAATAAT	3720
35	ACTATTTCTT TCGTTATAGA TAATTTACAG TTGTTTACGT AATACTTCAT CGTATTCTAA	3780
	GATACTTTTA CGCGCGTCGA AGTTATTACC TTCTACACGT TTTTGTGCTG ATTCTACAGC	3840
	TCTTGATACC ATTTTGTATT CAATTGGTGT AGAGTCATCT AAACCTAGTC GGCTCATCAT	3900
40	TTTCTGTAAA CGTTCAGAAC CAAAACGAAT CATTAAITCA TCTTGTAATG ATAAATAGAA	3960
	GCGACTATCC CCTTTATCAC CTTGACGTCC AGAACGACCA CGTAACTGGT CATCAATACG	4020
45	ACGAGATTCA TGTCGCTCTG TACCTATTAC TGCTAAACCG CCTAATTCCT CTACGCCTTC	4080
	ACCTAATTG ATATCTGTAC CACGACCAGC CATGTTAGTG GCAATAGTAA CGGCACCTTT	4140
	TTGTCCAGCG CCTGCAACAA TTTCAGCTTC ACGTTCATGA TTTTTCGCAT TTAACACATC	4200
50	ATGACGGATA CCACGTTTTT TAAGTAAATT TGAAATATAT TCAGAAGTCT CAACTGCAAC	4260
	AGTACCTAAT AGCACTGGTT GCCCTGCCTT GTGTTTTTCA ACAACATCTT CTACTACTGC	4320

55

## EP 0 786 519 A2

	TTTATTTGTC GGAATTTGAG TTACTGTCAT GTTATAAATA TTTCTAAATT CTTCTTCTTC	4440
	AGTTTTAGCT GTACCTGTCA TACCCGCAAG TTTATTGTAC ATTCTGAAAT AGTTTTGGAA	4500
5	TGTAATAGAC GCCATAGTTT TAGATTCAAT TTGAATTGTA ACGCCTTCCT TCGCTTCAAT	4560
	AGCTTGGTGT AAACCTTCCG AGAAACGACG GCCTGGCATT GTACGTCCTG TAAATTGATC	4620
	GACAATTAAT ACTTCGCCAT CAACAACCAT ATAGTCTACG TCACGTTGTA ATGTAACGTG	4680
10	CGCACGTAAA GCTGTGTTGA TATGACTAAT AACATCAACA TTTTGTACAT CATATAAGTT	4740
	TTCAACTTTG AACATACGTT CAGCTTTATC CGCACCTTGT TCTGTTAAAT GTACAGCTTT	4800
	CGTTTTTTCA TCGTATTTAT AATCTTCGTC CTGTTTTAAC ATTTTCGCAA AAACATTTGC	4860
15	TTGTGTATAA AGTGACGTTG ACTTTTCAGC TTCACCAGAA ATAATTAATG GCGTACGTGC	4920
	CTCGTCGATT AAAATTGAGT CAACCTCATC AATGATTGCA AAATGTAATG GACGCATTAC	4980
	TCTATCTTCA GAATAATTCA CCATGTTATC TCGTAAGTAA TCAAAACCTA GCTCATTATT	5040
20	AGTACTGTAA GTAATGTCTT GTGCGTATGC TTCACGTTTT TCTTCTGTCTG TCTTACTGTT	5100
	TAAGTTTAAT CCGACAGTCA AACCTAAGAA GTTATATAAC TCAGCCATTT CTTCACTTTG	5160
25	AACACTTGAT AAGTATTCAT TGA CTGTAAT AACGTGAACA CCTCTACCAG CTAATGCATT	5220
	TAAGTATGTT GGCATTGTCTG CTGTTAATGT TTTACCTTCA CCTGTTCTCA TCTCAGCGAT	5280
	ATCACCTTTA TGAATTGCAA TACCACCCAT AATTTGAACT TTATATGGTG TCATATTGAA	5340
30	TACACGTTTA GAGCCTTCTC TAACAAGTGC ATATGCTTCT GGTAAAATTT TATCTAAATA	5400
	ATCATTTTGC TTTTGTGACAT TATCAATGTC AGCTAATTCT GTTTGGAATT GTTTCGTTTT	5460
	ATTACGAATT TCTTCATCAG TTAAAATTGC CGTTTTTTCT TCTAAAGCGA TTACTTTATC	5520
35	AGCAAGTTTA CCTAACTGTT TAATTTCTTT ATTATTGCCA TCAAGAATTT TTGATAAAAA	5580
	TCCCATTTCG TTCGCTCCTT TAGCTAAAAA ACTGTTTGGC CTACAACAAT ATATCTTATC	5640
	ATTTATAGTT AGAAAATTAT ACTTATTTAC TCATTTGTAG AATCAATATA AATATATTTA	5700
40	TGACATACTT CATTACATT CTGTTGTCAA CAAGTTTATC ACTAATAAAT ATATTCTCAA	5760
	TACGCAATTA TACTTCCTAA TAAATTATAT TATAAATATT TTACGATTTT CGACTCGGAC	5820
	TATACAATAG ACTGACATAC TATTATTAAC TTAACATTCA AATATATACA TCCATTAACA	5880
45	TTAGCATAGT CACTATGTTT CATTCAACAA ATTACATTAT CGAACTATGA AATAGTCATA	5940
	ATTTGCTTTT GGAGTATAAA AAAGCACTTG TGCAAAAACA CAAGTGCTTT AACTTAATT	6000
50	TATTGTTTAC TAGTTTGAAT CAAGCCATAT TTACCGTCTT TACGGCGGTA AACGATACTT	6060
	GTTCCATCAG TTTCTCTGTC TGTGAATACA AAGAAGTCAT GACCTAATAG ATTCATTTGT	6120

55



ATCTCGTTAT CATCGTAAGC GTCATTATCA ACTTGTGTTT CTTGCATTTT TGTAATTTCG 6240  
 GCAACAAACA CTTCTTGATC TCCTCGATCA CGGCTCTTAC GATTAATACG TGTTTTATAT 6300  
 5 TTTGGAACCT GTCTTTCAAG TTTATTATTA ATTAAATCAA TACCTGCGTA TAAATCATCG 6360  
 TTTGCTCTT CAGCTCTTAA CGTAACATTT TTCAATGGAA TTGTTACTTC AATTTTAGTA 6420  
 GCTGAATTTG AATAAGTTTT AACTTTAACA TGCGCCACTG CATTGGGTAC GTCATTAAAA 6480  
 10 TAACGTTCCA ACTTACCAAT TTTTTCCKCA ATATAGTTGC GAATAGCATC TGTGATAGTG 6540  
 AGGTTATCTC CATGAATTTT AAATCTAATC ATAGTAAATC TCTCCTTAAA CCTCTTTATh 6600  
 GGAACCTChT TATTATATTT AACATTTTAA CGCCAATCGT GCAAA 6645  
 15

(2) INFORMATION FOR SEQ ID NO: 260:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7430 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

CAGTTCAGC ACATCTATTG GGGATCAACA AACTAGGGAA AATGCTAATT ATCAACGTGA 60  
 AAACGGTGTT GACGAACAGC AACATACTGA AAATTTAACT AAGAACTTGC ATAATGATAA 120  
 30 AACAAATATCA GAAGAAAATC ATCGTAAAC AGATGATTTG AATAAAGATC AACTAAAGGA 180  
 TGATAAAAAA TCATCGCTTA ATAATAAAAA TATTCAACGT GATACAACAA AAAATAACAA 240  
 TGCTAATCCT AGCGATGTAA ATCAAGGGTT AGAACAGGCT ATTAATGATG GTAAACAAAG 300  
 35 TAAAGTGGCG TCACAGCAAC AGTCAAAAGA GGCAGATAAT AGTCAAGATT CAAACGCTAA 360  
 TAACATCTA CCTTCACAAA GTCGAATAAA GGAAGCACCA TCATTAAATA AGTTAGATCA 420  
 AACAAATCAA CGAGAAATTG TTAATGAGAC AGAAATAGAG AAAGTACAAC CACAACAAAA 480  
 40 TAATCAAGCG AATGATAAAA TTAATACTA CAATTTTAAC AATGAACAAG AAGTGAAACC 540  
 TCAAAAAGAC GAAAAAACAC TATCAGTTTC AGATTTAAAA AACAAATCAA AATCACCAGT 600  
 AGAACCAACA AAGGACAATG ACAAGAAAAA TGGATTAAAT TTATTAAAAA GTAGTGCACT 660  
 45 AGCAACGTTA CCAACAAAG GGACAAAGGA ACTTACTGCA AAAGCGAAAG ATGATCAAAC 720  
 GAATAAGTT GCCAAACAAG GGCAGTATAA AAATCAGGAT CCTATCGTTT TAGTGCATGG 780  
 50 TTTCAATGGG TTTACAGATG ATATTAATCC TTCAGTGTTA GCTCATTATT GGGGCGGTAA 840  
 TAAATGAAC ATTCGCCAAG ATTTAGAAGA AAATGGTTAC AAAGCTTATG AAGCAAGTAT 900

55

	TCGTGTAGAT TATGGTGCAG CACATGCAGC AAAATATGGA CATGAACGTT ATGGAAAAAC	1020
	ATACGAaGGA ATTTACAAAG ACTGGAAACC AGGACAGAAG GTACACCTAG TTGGACATAG	1080
5	TATGGGCGGT CAAACGATAC GTCAACTAGA AGAATTACTG CGTAATGGTA ATCGTGAAGA	1140
	AATAGAGTAT CAAAAGAAAC ATGGTGGcGA AATTTCTCCA CTATTCAAAG GTAATCATGA	1200
	CAATATGATT TCATCAATTA CTACTTTAGG AACACCACAT AATGGTACAC ACGCATCAGA	1260
10	TTTAGCTGGT AATGAAGCTT TAGTGAGACA AATCGTATTT GATATCGGTA AAATGTTTGG	1320
	TAATAAAAAAT TCAAGAGTAG ACTTCGGGTT GGCTCAATGG GGTCTAAAAC AGAAGCCAAA	1380
	TGAATCATAT ATTGATTATG TCAAACGCGT TAAACAATCT AATTATGGA AATCAAAAGA	1440
15	TAATGGATTT TACGATCTGA CGCGTGAGGG TGCAaCAGAT TTAAATCGTA AAACGTCGTT	1500
	GAACCCTAAC ATTGTGTATA AAACATACAC TGGTGAAGCA ACGCACAAAG CATTAAATAG	1560
20	CGATAGACAA AAAGCAGACT TAAATATGTT TTTCCCATTT GTGATTACTG GTAACCTAAT	1620
	CGGTAAAGCT ACTGAAAAAG AATGGCGAGA AAACGATGGT TTAGTATCCG TTATTTCTTC	1680
	TCAACATCCA TTTAATCAAG CTTATACAAA AGCGACAGAT AAAATTCAAA AAGGCATTTG	1740
25	GCAAGTGACG CCTACAAAAC ATGATTGGGA TCATGTTGAC TTTGTAGGAC AAGACAGTTC	1800
	TGATACAGTG CGCACAAGAG AAGAATTACA AGATTTTGG CATCATTTAG CAGACGATTT	1860
	AGTGAAAAC T GAAAAGCTGA CTGATACTAA GCAAGCATAA TTTATAAAGT AAAGGGAGGA	1920
30	ATTAATAATG ACTGCAGACT TCTTTCAATT AATCGGATCA TTATTTAGAA TTCTAAAAGA	1980
	ATTATTCAAG TAAAACATTG GCGAGGCCCC AACATAAAGA ATTTGAAAA GAAATTCTAC	2040
	AAACAATGCA AGTTGGCGGG GCCCCAACAA AGAAGCTGGC GGAAAGTCAG CTTACAATAA	2100
35	TGTGCAAGTT GGCGGGGCCC CAACATAGAA GCTGGCGGAA AGTCAGGTTA CAATAATGTG	2160
	CAAGTTGGGG TGGGACGACG AAATAAATTT TCGGAAAATA TCATTTCTGT CCCACTCCCA	2220
	TTGGCATTTA CGAAGTTTAA ATGTGCAATT AGAATATATG TATAACAATA TTAAACACGC	2280
40	GGTAAAACGA AGTCAGTCAA TTCAACTGA TTTGCCCCAC CGCGTGT TTTT TAACATAGCT	2340
	TAATAATTAA TAAGCATTA TGTTCATTT ATATGGTTGT TTTCCAATAA TAAACCTAAA	2400
	GATATAGAAT TCACGCAATA TCATGCCGAC ACCTATACAT AATCCTAAAA TGAATAGTAG	2460
45	TGATATCGCT AGAAAGACCA TTGTATTATC CTCAAATATA TTTGTATATG CAAACAATGA	2520
	GTCTAGAATG ATTGGATGTA ATAAATAAAT AAAGAATGAG AAAGCACTAA TCATTTGAAT	2580
50	CGTATTAAAT AACATTGTTT TAAAATGCGT GCAAATACCC AAGATAACAA TAAACATAAT	2640
	ACTATTATAT GGTGTTAATG AATATGAAAA GCTGGTAACG TTCCAATAGT CTCCaTTTGT	2700

55

	TCTAAGAAAT TTAATACACG TTCGTAGTTA TAACCCATAT ATGCACCTAA GAAGAAATAA	2820
	AAAATCCATC CGAATATTAT AGTATTTTCA CTTAATGGAT AATAGTGTAG CACGGTATCG	2880
5	TGAAACGCTG TGTGTTCGT AAAGTAATAT AAAAATGATT GCTGTAAAAT AAAAGATAAC	2940
	AATAATAATA TTTTACTGTT GAATAGGTTA TAGTTAATTT TAAAAATGAT ATAACTCAA	3000
	ATAAAGAATT GCATGATAAC AACGATAAAA TAGCCATACC ATTGACCTAA TAGGACATTT	3060
10	TCAATGAATT GTTTATTGAA ACTTGAATCT GTTAATAATG ATTCACTATA ACTGTAAAAC	3120
	AATCCCATTA ATATGTAAGG AATAAGTATA TATTTTACGC GTGTAGTTAA GTATCTATAG	3180
	GTGACTTTTT GGTAATTCAA GGTGTGAGT AACTGTGACA AGATAATAAA GCAAGGTGTA	3240
15	CCAAAAATCA CAATATTACG AATGTAAAAT TGTAACACTA AGGATCCACC CTCCATATTT	3300
	TCATGTTTTA AAGTAATTG TGTAAGTAAA TGTGTGATAA TAATAATTGC ACATATAATA	3360
	GCACGTAAAT ATACGAGTTC AAGTCTAATC TTTTTCATGG AATCCGTCCC ATCTCTTAAT	3420
20	TAAATGCTCA AAAGCATCAT CACTAATTAA TATTCTAGGG ATGTAATAAT CATTGGAGTT	3480
	CGGAGTGACT GCTTTTTCTT CTAATGAAAA ACCGTATTTT AACCAGCTT TTTTGATTAC	3540
25	CGGTAATTIA TCGTCATTCA TCAAGCCATA AGGATAGGCT ATAGTTTCTT GCGACTTTTT	3600
	AAAGTTTTTA GTTAGATATT TTTCACTTTT GTTTAAATCT TTTATGATTG TAGCTTCAGA	3660
	AGCTTTCATT AATTTTGACT TATTATTTTT AGATAAGTTA TGCAAATCGT GGGTATGTGT	3720
30	TTCAAATTCC CATAACCCAG TTTTATACAT TTCTTTTAGT TCTTTTTTAC TAATCATATC	3780
	GAGGTGTGTA AAGTTTTCTT CCCCAACATG ACCTGTGATA ATAAACCCAG TTGCCGGTAT	3840
	TTTATATTTT TTTAAGATTG GATAAGCATT TTCATAAATA GTTTCATCCA TATCATCAA	3900
35	GTAAATCCAT ACACTTCGTT TTGGAACTT ACCTTTTTTC TTGTAATATA AAAATTCTTT	3960
	CAAGTTAAA AATTTAGCAT CATGTGATTT TAGCCATTTT ATTTGAGATT CAAATTGTGA	4020
	TTGACTAACA CTATAATTTT TAATTTCTTT ACTACTAGAA AAGAAGTAAA TAAAATTATT	4080
40	CAGAAAATTC GCTTTTCTTA CACGGTGATA ATTTAATGCC AGAGCACTAT TTTCTTTATA	4140
	TTTCAGTTTT TTAGGTGAAT CGTCATCTGC ATTTGCAATA TGATGACCAT CCAGTGTGCT	4200
	TACAGGCAAT ATGATCAAGA TACTCAACAC TAAAATTATA AATTTTCTAT ACTTCACGAT	4260
45	TCTCTTCCTC TCTGCCATTT TTGAATCAAT ATGCTAATTG TAAAAATAC AAAAATGATA	4320
	ATCGCGAAAA TGCCCATAGT TTCAAATATA TCTAAAATTT CAGTATTTTC AATGTTTAAA	4380
50	GCAACACGTA TTGTATTGAT ACTTTCGTCA TGAATTTCAA ATATAGTACC AATATAAACG	4440
	AGTAGAACAA CTAAACAATA TATCCAAAAG ACACAAGATA TAGCGATAAG TGCTGTTTCT	4500

55

	TTGCGTAACC	ACCTTTCTTA	CGTTTTAATG	CTTTTGAAA	TGCGACAAGA	ACTACTGCTG	4620
	CGTTAATAAT	CCAGTATACT	GTCGGATACC	AACCTACAAA	TATGAGTCCA	GCCATATTCT	4680
5	TTTTCTCGTA	GCGACTATCA	ATAAAGAGTG	CGACTGTAAA	TTGAATAACG	TTTATAAAAG	4740
	TCATAGTAAA	TGATGATAGT	AGAAATATTG	AAAAACTATA	TGTCATAAAT	GTATAGTCTA	4800
	AGAAGTTTGC	TGTTATGAAC	AAATAGCCTA	AATATAGAAG	CACTATATAT	ACCCATAAAA	4860
10	TCGAGATGAT	TTGCTCAAAC	ATCAAAATAT	ATAAAGGAAA	CCTTTTCGTT	TTCAATTGTGC	4920
	TAAAAAAGTC	TCGTAGTAAT	ACTTCGTGTC	CCCCTTGAGC	CCATCTCACG	CGTTGCTTCC	4980
	AAAGACCTCC	CAATGTTTCT	GGAACCAACA	TCCAACACAT	GGCAAGCGGT	TCATACTTAA	5040
15	TACGATATCC	ACGTAAATGC	AATTTCCAAG	AAACTGCAAT	ATCTTCGGTA	ATCATATCAG	5100
	TATCCCAGTA	GCCAACGTCG	ACAACTGCAC	TTTTTTTAAA	TAGAGTGAAG	ACACCCGAAA	5160
20	TAGTATTGAC	TGCGCCAGCA	AGTGTCTGAC	TtCGCTTAAT	ACAGCCAATT	AAACTTGCAT	5220
	ATTCTATCGT	TTGAATTTTA	CCTAAAATAG	AACTCTTATT	TCGAATTCTA	GGATTACCTG	5280
	TAACTGCACC	AAGTTTGGGA	TCATGTTTGA	AATTCTCAAT	CATATAATAT	GGTGCATCTT	5340
25	GATCAACGAT	AGTATCTGCA	TCCAAGCACA	TTACATAATC	ATATGAAGCC	TGTTTAATGC	5400
	CTTGATTGAG	TGCGTTGGCT	TTACCTCTGT	TTTCTTGTA	ATCGACGAAA	ATAAAGTCAT	5460
	TATTTTCTTT	GATTTTATAG	ATGAGTTCTG	CTGTATTATC	TGAACCTCCA	TCATTAATGA	5520
30	TAATAATTTC	TTTCTTCTCG	TATTTGAGTG	CAAGAACATT	AGACAACGTA	TCTTCAAYCG	5580
	TTTCACTTTC	GTTATAACAG	GCAAGTAAAA	ATGTAATGCC	TTCTAATTCA	TCCACATTTA	5640
	TGTCAGGCTT	CTTGTTCAAT	GAATATCTAA	TTTCTCTGGT	AAAATAGAAA	TAAATTGAAC	5700
35	CGACAATCCA	GTAAATAGAC	ATAAATACAG	GATAAAAAAG	CAAAAAGTTA	AAAAATTGCA	5760
	ATTCTTTTAC	CTACCTTTCTG	TTAGTTAGGT	TGTAAGCCAT	ATGGTAATTG	ATAGTATTTT	5820
	AATTTGCAAT	AGATTGTTGT	TATAATTAAA	CGGAAATATT	TGTAATTGCA	ACTTAATTTT	5880
40	CCTGTAACAT	AGTGTGATTA	ATTTTCAGTA	GGGGGTATATA	AAAATTGAAG	GATAAGATTA	5940
	TTGATAACGC	AATAACCTTA	TTTTCAGAGA	AGGGGTATGA	CGGTACAACA	CTTGATGATA	6000
	TAGCTAAAAG	TGTAAATATA	AAGAAAGCGA	GTTTATATTA	CCATTTTGAC	TCGAAAAAAA	6060
45	GTATTTACGA	ACAAAGTGTT	AAATGTTGTT	TTGATTACCT	TAATAATATT	ATTATGATGA	6120
	ATCAAAATAA	ATCGAACTAT	TCAATTGATG	CTTTATATCA	ATTCTTATTT	GAGTTTATTT	6180
50	TCGACATCGA	AGAAAGGTAT	ATTAGAATGT	ACGTTCAATT	ATCTAATACG	CCTGAGGAAT	6240
	TTTCTGGAAA	TATTTACGGA	CAAATACAAG	ATTTAAATCA	ATCATTAAAGT	AAAGAGATAG	6300

55

TGCTGTTTCT TGAAAGTTGG TATTTGAAAG CATCCTTTTC GCAAAAATTT GGAGCAGTGG 6420  
 AAGAAAGTAA AAGTCAATTC AAAGATGAAG TGTATTCGCT ACTAAATATA TTTTGAAGA 6480  
 5 AATAATTTTT GTTACTAGTT TGTAATAATT AACTTACTTT TGTAACAAAA GACATGAGAT 6540  
 TATTTTTTTA AATCTATATA AAGTTGACAA TACAAATCGA TATTGAGAAT ATTAAGATGT 6600  
 ATATGAATTT TATAAATTAA ATGCAATACA TTAATATAAA TATCAATTGT TGCAAAATAC 6660  
 10 GATTTGTTCA ATGATTTGAT AATATTATTC TTTATATTTG TGAATGGTTA AGTTTGTCTT 6720  
 TGAACATATT ATAAAAGTGT AATGTTCCCC TGAAAAGAAT AAGTTGTCAT CTAATTACAG 6780  
 15 GAAATCCGCA TAAATTAGAT GAAATGGAAG GTAATAAGTA ATAATTTATT GATAAGCGCC 6840  
 TATGTGATGG TAAATCATGA CATAGGCGCT TTTTTTTATA AGTTAAAAAT GTAAATAAAA 6900  
 ATTATATAAA TTACCCACAT CTTTTTAAAA GGTGTGGGCT TTATTATCAT TAACCCAACT 6960  
 20 CACAGTGACG GGTTACGCAA GGTATTGAAT TACCGAGTAC GGGCACGCTC GGTGTTGTAA 7020  
 AGAGCAAATA ATCAAGTAAT GATGATGCTT CTAATCGATT ATAAGAAAGC CATGATAGAG 7080  
 TACGATGGTA TCTAGTTTTA TTATTAATAG GTTTGGATAT TTAAAGTTGG ACAATATTAT 7140  
 25 ATCTTGTGCA AAAATATAAA TAAGTTATAC ATAATGGTAG AGAATCATGA TATAATTTTA 7200  
 AACGATAAAA TATTTATATA AATAATTAGA GaAAATGTAG TTGTGTATGT yTTGTGGtCG 7260  
 TTAAACTAGA TATAATTGTC CGATTTATAA AACATACATA ATGAATACaA TGATTGATTa 7320  
 30 TGTGGAGGAA ACCATGAmAG AmAAGTTTgA TTTAGTAAAA CTATTAAATA TTCTAAAGAA 7380  
 GAATATTAAA TTATTGCTTA TTTTACCGGC AATATGTCTT GTAGTAAGTG 7430

35 (2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4082 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

ATTGTTACTC ATTATTTTTT CATATTCACA CAAATGATCT TGTTTATATT TAGCTAATTG 60  
 ATTTTATCT AGCATTTTAT CCTCCTGCTG AGTTTGTAAC CTTTAATAAT TTATTTTCTA 120  
 50 TAAAAACTTA GTATTCCAGT TGCTTATTAT ATCATTGATG AAAGGCTGAA ATAAACATA 180  
 AACTGTTCCG ACCATTAACG CTGTAGCTAA AGATAAGTCT ACAAGTCCAC CTGTTTTAAA 240  
 TTGAATCGGT GTCTTCACAT TAAACGGTAA gGAKnAAAAT AATTTACACG CTTTGGTGT 300

55

	AGGCGTCTGA ATAATTTGCA ATAAAAATGC TATGATTGCG ATAAATAATA TTGAATGCGT	420
	AAAGGTTTCGA TGTCCAAAGA TCAATCTCAC AAAAAAATA ATTACCTTAA ACCTTCTGCC	480
5	AATCTTACTT TGAGTGTGAC ATATATCGGG TAATAAGCTA GCTAGAGTTG CTAGAATGAT	540
	AACCGTAACC GACGAAAAAA TATCCGTTTG AAAATATTGT GTTGTTAGCG CTCCAACGAG	600
	CATGCCGCAT GAAGCATGTG TTTTACCTGT CATATTTGTT CTCCTTTAAT ACTCACATTT	660
10	TACCACATCC CTAACAAAAA CACGAACATA TTTTCGGGTT AAAATTCATT AGTATGACAC	720
	AATTTAAAAA AGTATCACAT AACTCTTGAA AACGATTACA AAATCGTTTA TGATGTATTT	780
15	ACAAAATATT TAAAGGATGT GTTGAATAA TGGCAATGAC AGTAAAAAAG GATAATAATG	840
	AAGTGCGTAT TCAATGGAGA GTTGCTGATA TCAAAATTCC TACAAGTGAA ATTAATAATA	900
	TTACACAAGA CCAAGATATT CATGCAGTTC CTAAATTAGA CAGCAAAGAT GTATCTAGAA	960
20	TCGGCTCAAC GTTTGGTAAA ACGAATCGCG TTATTATCGA TACTGAAGAC CACGAATACA	1020
	TTATTTATAC TCAAAATGAT CAAAAGGTTT ACAAATGAAT AACTAAATAA ATTGTATAAA	1080
	AaAATCATT C ATGGTGAGGG CTTCATGAAT GATTTTTTTA ATTGATTCAA CACCCAGCAT	1140
25	AAACAAATAC AAAAGGACAA CTGTTCCTCAT AATTTTAACA GTTGTCTTTT TTCACATATA	1200
	TTTATAACAA AAGATGTGCC ATCAAAGAAA TAATTGGTAG TGTAATGATT GTTCTAATCA	1260
30	AGAAAATCAT AAACAATTTG CCGATGCTTA CAGGAATCTT CGAACCAAGT ATGACGCCAC	1320
	CTACTTCAGA CAAGTATATT AACTGCGATA TACTAAGTGC CCAATAACA AAACGAGTTA	1380
	TATCATTTTG TACACCTTCA ATTAATATAG AAGGTAAAAA CATATCGGCA AAACCGATAA	1440
35	TAATCGTTTG AGAAGCCTGT GCCGCTTCAG GTATTTGCAT TAACTCTAAA AATGGAACAA	1500
	AAGGTTTACC CAATATGACA AAAAAGGGCG TGTAGTTGCG AATAATGGTA GCAATAGTAC	1560
	CAATACTCAT TACTACAGGC AAAATAACAA ACCACATATC AATGACTGTT TTTAATCCTG	1620
40	ACTTAAAAAA GTCAATAACG CCCGGTGCTT TAATACCTAC TTCTGTTGCA GTATCAAAGC	1680
	CATGTCTCAA TGCCGTCTTT CCTTCTGGCA ATGCCTCAGT ACGCGCACTT TCAGGTACCT	1740
	CCTTAGCATA CTCATCAGGA ATTTTATTTA AAGGCCAAAT TCTTGGCATA ATGACTGCTG	1800
45	CAACGAGGCA GGATACTATC ACTGATAAAT AGAAAGCAAA AAATTGATTT TGCATGTGCA	1860
	CTGTTTCAGC AACTACAATT GCAAAGGTGA TAGAACTAC ACTAAATGTC GTTGAAATAA	1920
50	CTGTTGCCTC ACGACGAGAA TAATATCCTT CACCATATTG TCTACTTGTA ATTAAGACAC	1980
	CAACAGTTCC GTCTCCAATA AATGATGCTA AATTATCTAC CGTCGAACGT CCTGGCAATG	2040
55	TAAATAAAGG TCTCATAACC GGTCTAAAAA TAGGACCCAA CATCTCTAAC AAACCGTATT	2100

	AACTTGAGAA CAATAATCCA CCCGTTTCAT CTGAGTAAAT AACCTTTGAA CCAATTCGTA	2220
	AAAATGTCAT CCATGCAAAA ACAACTGCTA ATATTGCTAA AATTAACCAA CCAATTCTAA	2280
5	CGTTAAAAGC ATTGTTTCATT AGCCCGTCAG GTTTC AATTT ATCTTTTAAA ATAGTTGAAC	2340
	AAATCAGAGT TATGATACCC GATAAAGTAA TTATCGTCAC AATTA AAAAT GGCATTACGC	2400
	CACCTAATAC ATCTTTAAGC ACGCCTGCTA AAAATGCCAC GGGCAACGTT GTTTGCTTCT	2460
10	GTCCATCTTG TTCGACTGGA ATTGGTACTA AAAATAATAA GATACCAATT AAAGACATCG	2520
	TAATAAACTT AAGTCTCCCA ATA ACTATCT CTTTCCTTGA AAAGCTATCC ATAAAATCAA	2580
15	TCCATTTCTC TATGTATTCTG TTTTAAGTAT ATACAGAATT CTATTCAGTT AACAAACATA	2640
	TTCTTATCA TTCTATCTTT CAAAATGTTT ATGTATGCAA AATAATGAAT AATTACAGTT	2700
	ATTAAATATA CGCTATTTCT TGTAATTTTT CAAGATGAAT TCAAAAAGG TTAAGTACAA	2760
20	TTACTGATTT CGTACTTAAC CTTTTTTAAA CTCTAATCAT ATGTTAGTTA TTTCATTCTT	2820
	CGTAATAATA TTAAGAAGTA TGGTGCACCG ATAATTGCAA TGATAACCCC AACAGGAATA	2880
	TCCAGTGGCG GATGAATGCC ACGGGCTAAA CCATCTCCAA ATGTTAACAA TATAGCACCA	2940
25	ATTAACCCCG ACATGATAAT AACGTGTAAT GTTTTATTTT CTATTAATTG TCTCGCAATA	3000
	TGAGGTGCAA TTAATCCTAA AAAGCTAATA CCACCGACAA CTGAAATTGC GGATCCTGCT	3060
	AATATTACTG CTAAAATTAA CAATAGCATT TTAATAGTTT TAACTTTTAA ACCGAGTGCG	3120
30	GTTGCAACAG CATCACCTAG ATTCAATACA TCTAATTGAT AACTCAATAA AATGATGATT	3180
	GGTATCGTTA TTA AAAACCA AGGTAATATA GTATAAATAT TCGACATATC ATGTCCATAT	3240
35	AGACTACCTG TCAACCAAAC AAGCGCTTTG TTTGCTTCCA GTGGATTTCT GATTAATAAG	3300
	AACTGCACAA TCGCCGTACA TATTGCGCCT ATTGCTAAAC CAATTAAGGC AAGCTTTGAA	3360
	CCTTTAACAT CATATTTTGA AATTAAAAAT GATAAAAATA AACTTACTGC AAAGGCACCT	3420
40	AAGAATGAAC CTATAGGTAA TACAAACAAT GGTGCTGTTG GAAAGGTCAT AATAATAATC	3480
	ACAGCAGCTA AACTGGCACC TTTAGAAATA CCTATAACAT CAGGTGAGGC TAACGGGTTT	3540
	CTTATTACAG CTTGTATAAT TGCACCTGAA ATAGCCAAGC TACTACCGAT AATAATACCA	3600
45	AGTAATGTTT TAGGTATACG ATACTCATTT AAAATAAAAT CATCTGTGT AAAGATTCCC	3660
	TTAATAGCAT CAATCGGATG AATCATGACA GACCCTACAC ATAAACTTAT GAATATACTC	3720
	ACAATTAAAA GGATTGTGAT TAACTATAA CGACGTATAA TTTTCGTTGT CATCATATTC	3780
50	TTTTACCCCC TTTAATCGTT ATAAATAAGA AGTAAAGTGC ACCTACGAAT GATGTAACAA	3840
	TCCCTACTGG TGATTCATAA GGATATGTAA TTAAACGACT TAATACATCT GATAGTAGTA	3900
55		

AGCGTTTGAC TATATGCGGT ACGATTAAGC CAACAAATCC AATTGGTCCT GCCACTGACA 4020  
 CCGACATACC TGTAAGAATA ATGACTAATA GTCCAATGAT AATTCTAACT TTATTTATAT 4080  
 TT 4082

(2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1145 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

TATTAGAAGG TCGTTCGGAT GAACAATTAA AAAATTTAGT TAGCGAAGTA ACTGACGCCG 60  
 TAGAAAAAAC AACGGGGGCA AATAGACAAG CAATTCACGT TGTATAGAA GAAATGAAAC 120  
 CAAACCATTA TGGTGTGGCT GCGTAAGAA AGTCAGATCA ATAATCTTC ATAAGATGCA 180  
 TGCCAATTAA TTCTTTGAAA ACGAACAAGG CGACTTCTAT CTGaGTATGA TAGAAATCGC 240  
 CTTGTTTATT TTTAATCTTC ATCTAAAAG TCTTAAATAG CTTGTTTATT TGTGTTTTTA 300  
 TTAATCTGTA ATGCACTACC ATCAGTATTT GTATTGACAT CyTCGTATGA GTTCTTGATT 360  
 GGCaCAGTCA ATGACTTAAC ATCTTTTTCA CCTCGGATAC CaAAACTCAA ACCTGTTTGG 420  
 AAAATCCCTG AATCAGGAAT GTTGTATTTC ACATAGCCTC TTAAATACC TGCAACTTTT 480  
 GGTAATTTAA CAACTGTTCT AAAATTAACC ATTTCTTTTT TCAATGTTTG CATCACTTGT 540  
 TGCTGACGTC GCACGCGTCC GAAGTCACCT TCAGGGTCGT GACGGAATCT TGCATAACCA 600  
 AGTAATTCTT TACCATTCAA CCTATGGTTA CCCTTTTTCA AAGATACACC AATATTTTTT 660  
 GACATATCTT TTTCGACATT AATTGGTACA CCTTCAGGCA TTAATTCATC AATCATTTTC 720  
 TCAATCCAG TAAATCAAC TACTGCATAA TATTCAGGAT TAATTCCTAA ATTTTATCA 780  
 AGTGTTTTTT TAAGTAGCTC TGGACCACCT AAAGCGTATG CTGAATTAAT TTTGTGTTTT 840  
 CCATATCCTG GAATATCTGC ATAAATATCA CGCATGACAG ACATCATTTT CATCTTTT 900  
 TTGATAAAGT CATATTGAAC AACCATGATA GAATCTGTTC TTGATTGTCC ACCTTGTGCT 960  
 TTATCTGCAC CGAGTACAAG AATAGAAATT nTACCATCAT TTTTACTGG TCCATTAAAT 1020  
 TGATGTACTT TAACATCTTT CGCATGTTTC TTGGCATATT CTACACCGCT ATTGTAACTA 1080  
 TGTACAATAT ATACAACTAA TGCCGATAAG TAAATTACA ACAATCAGAA GAATGATAGG 1140  
 TAATT 1145



## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7075 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

10 TATGGCTCAT CATTAAATGCA CGTATCGGGT AGCGTTTACC ATTGATAAGT GCTTCATGTT 60  
 TAGCACGAGT TCTTAAAATT CCATCGCCAT AACCGATATC AACTACAGCT AATTTTGTAT 120  
 15 TGTTTTTAGT CACTTCAAAG GCAAAGCTAT AACCGCAATA ATCACCAGCT TGTACTTCGC 180  
 GCACTTGAAT AACATGTGCT TTAAAGTTA ATGACTGAAC TATATCATGT TGATTCAGTG 240  
 AACTATATGG TCTTGAACCG TATAACGCAA TACCTACAG CGCATGTGTA TGGTGGGGTA 300  
 20 GTAATayyyg TCCTTCCCGA TAAAACTCG CACTATTTTG AGCATGGATT AGGTCGAACT 360  
 GATAACCTTC AGATAAAAGT GCTTCAACAA TTTCCATCCA TTGTGAACGT TCAACATTAT 420  
 AAtCTGACAC ATCGAATTCA TCAGCATATC CAAATGGGk CCATAAACCA CTAATAATCA 480  
 25 TTTTTCGATT TTGATTATGA TGGTGATCTT TCAATACTTC TTTAATTCG TTTAAATCTT 540  
 TAAATCCAGA CCGATGTAAT AAATTTTCAA ATTCTAAGTG AACATGAATA CCAGCTAAAT 600  
 CATTTTTATG GTTATAGTAA TATGTCAACG ACGGCAAAGT CATGTGTATT TGATGTTTAC 660  
 30 GGACTAAATC AAACGTAAT ACTGCATTCA TAAAAAGAT TGTTGCATCT GGAGCAAGTT 720  
 GTCTAATTTG AATTGCTTCT CGTAGTGATG TTGTGCTAAA TGTATCTATA CCTGCATGGA 780  
 TAAACTGAGT TACAGCAAAT TCTAGGTCAT AGTGATATGC ATTaTTTTTA ACAACTGCCA 840  
 35 TTAATGGCTG ATTGTTTTTG ACTGTGATTG CATTTTGTAA AAATATTTTC TTATTTACAG 900  
 ACCATGTTGC TGTCAATGTA TTACACCTCT TTGTAATTAT TTAATAAATT TTCGTAAAAA 960  
 40 TTAACCACGT TTATTAACAC TTTTTCATCA AAATTTAAAT GTGATGTGTG CAAACCAGTT 1020  
 ACAAACCTT TATCTTCATT TCGTGTTCTT ATAAAAACAA AGTAAGCTGG AGCTAGTTGT 1080  
 TGACCATAAA AACTAAAATC TTCCCCAAAT AAGAATGGCG TTGGTTTGTC ATAGACATTT 1140  
 45 AAATCAGCTT TTATTAAGGC GTCCTCTATT TGAGTACGTA ATTTCCGACT ATTGATTGTA 1200  
 GGGGGATAAC CTTCTGCAAA TTAACTTCA CAATCTACAT TAAACAGAAG CTTGACACTT 1260  
 TCTGCTATCT TGTGCATTTG ATTTTAAACG ATTGTTAAAT CATCAATATC ATATGTACGA 1320  
 50 ATAGTACCTT CTAAATAGCC ATTACTTGGT ACAGTGTTAA TCGCTTCACC AGCTTTAAAA 1380  
 TGACCAATAT GAACAATATT TCGTTTCAAA CCGTTAAGGT GAAATTGTTG AATTTGTGAT 1440

55

	ACATGACTTG ACAGGCCTGT TAAGAAAAAG CGATACTCTG TTGCGCTGGC CGTAATTTCT	1560
5	TCATCTCTTA TCACTGCAAT GCCTTCATCA GCAAATGGGT TAACATGAAT ACCAAATACC	1620
	GCTTCAATTG GATACTTATC AAAGGCACCG GCTTTTATTA ATCGATTTGC ACCGCCACCA	1680
	GTTTCTTCTG CAGGTTGGAA AATGAAAACG ACATTTTGCG GTAATTGACC TGCATCTTGC	1740
10	ATGTCTTTGC AACGTTGTAC AAAAAGCATT AATGCAGTTG TATGACCATC ATGTCCACAA	1800
	GCATGCATCA CATGATCAGA TTGACTGCGA TAAGGCACAT CATTTTCCTC TAAAATAGGT	1860
	AACGCATCAA TATCAGCTCT ATACGCTATC GTATGTGAGC CATTACCTTC TAAGTATGCA	1920
15	ATGACGCCAG TTTCCAATGG GCAATCGTAT TTAATATTTA AACTATCTAA AAACGCTTTA	1980
	ATATAAGCAG TTGTTTCAAA TTCATGTAAG CTTAATTCAG GATGTTGATG TAAATGACGG	2040
	CGATGTTTCG TAACAAATTC TAATTCATTC ATAATTATCA ATCCTTTGTG TTAAATTACT	2100
20	ATATAAATAG TGTAACGLAT TTCGAAATTT GTGATCATAA GTTTATTCAA TGCTAAACAA	2160
	TAAGGTTGAG ACATAATCGT ATCTCAACCT TGAAATTATT ATACGTTGAC GTCAGTAGTC	2220
	ATTGAGTTTT CTTAATGCTG CTACAATCTC TTTTITAGTA TCTTGTAATT CAGAAGCTTG	2280
25	CTTAATCACT TTTGCAGGTG TACCAGCAAC AACTGCACCA GCTGGTACAT CTTGTGTCAC	2340
	AATCGCGCCA GCTGCAACAA TAGCACCTTT ACCAACACGT ACACCTTCTA AAATAACTGC	2400
30	ATTTGCACCG ATTAATACAT CATCCTCGAT TATAACCGGT GAAGCACTAG GGGGTTCAAT	2460
	CACACCTGCT AATACTGCGC CAGCCCCAC ATGTACATTT TTACCAGTTG TAGCAGGACC	2520
	ACCGAGAGTA GCATTCATAT CAATCATTTG ACCTTCGCCA ACGACTGCGC CAATATTAAT	2580
35	TGTTGCGCCC ATCATAACGA CAGCACCATC TTCAATAATG GCTTGTTCTC TAATAAACGC	2640
	ACCTGGTTCA ATTGTTGCAT TCGTATTGTG TAAGTCTTTT AATGGAATAG CAGAATTGCG	2700
	ACGATCCATT TCAATTTCTA TATCTTCGAA TTGACTACCA TATGCTTCGT AAAAAGGTTT	2760
40	CCAATCATCC GCTTCACAAA AGATTACTTT AGATTGTTCT GAACCAAATA CTTTAAACT	2820
	TTCTGGATAT GTGATGCCTT CAAAATTACC ATTTAAATAT ACTTTTATTG GTGTAGACTT	2880
45	TTTAGCATCA CTTATATATT GAATAATTC TTCAGCTGTT AAATGTTGTA CCATAAAATA	2940
	ATCGATCTCC TTTAATATGT TTATAAGTTG TCAAACGTAT AAAAGCCGTT TGGTTTATTA	3000
	ACTAAGCGTT CTGCTGCTTG TATTGCACCA TTCGCAAAA TATCTTTTGA TTGTGCACGA	3060
50	TGCGTGATTT GAATCGTTTC ATCAGTGCCA GCAAATAGAA CTTCATGTTC ACCGACAATC	3120
	GTACCTCCAC GAATAGAATG TATACCAATA TCTTGTTGGT GGCCTTTTTC ATTTAATTCA	3180
	TGTCTATCAT ACACAGGTGT TACATTTTCT TTCAAAGATA CGATCACATC ATACAATTTT	3240

55

	TCGAAATCAT CAAGTAGGGG AACAGCAGCT GCTAAAATTT TAGTCAATGC ATGAACGCCA	3360
	TAACTCATGT TCGCGCTGAA AAACACAGGC ATATTTTGAC TCAATTCATC TAACTTATTA	3420
5	AGTAGTTTTT CTTTCTCGnC CAGTgTTGCC ACAACTAATG GCAAATGAAA ATCTTCATCT	3480
	AATAAAGGGA AAAGCAGATT TGGATTTGAA AAATCTATTG CAACATCGGC ACCTTTAACA	3540
10	TCTGCAATAT GTTGATATTG TTGATATGGC GTTGTGCTT TCGGTGTATT TTCAATGACC	3600
	CCAACGATTT CATGTCCTTT TTCTTCTGCT AATCTAGCAA CGCGTTGATT CATTGCGCCA	3660
	TAGCCAATTA GTAATATTTT CACTCATTTT CACCCGCTTT AAATGTGTCA TATGTTTCAC	3720
15	GAAGCACTTT AGTATCTGTA TCTTCTAGGC TAACCAATGG TAGACGTAAT TCATAATTTT	3780
	CAAATCCTAA ATAACCTGTT AGAGCTTTAA TAGGAATTGG GTTAATATCA ACTGATAAAG	3840
	CTGATAACAG TGTGCCGATT GGTTTAAATT GATCTTGAAT ATCTAATCCA CTTTGTGAG	3900
20	CATCGTATAA CGCTTGAAAT TCTTTAGGAA TGACATTGGC AATAACAGAG ATAACCCCTT	3960
	GACCGCCACG TTGATAGTAT TCGACGACGT TGTATCATT GCCACTATAT AATGCAAATG	4020
	AATTGTATC AATGCGCTTT TTCATTCTT CTAAATACTC AAAATCATTC GTAGCATCTT	4080
25	TTAAAGCAAC TATATAAGGA TGTTGACTTA ATATTTCTAC AGTTTCTGGT TCAATTGTCA	4140
	TGTTCTGTTCT TGAAGGAACA TTGTACAGCA CGACTGGTAA TTTCACAGCA TCTGCAATCG	4200
30	CTTCAAAGTG TTTGACTAAA CCACGTTGGT TCGTTTTGTT GTAGTAGGGC GTAATTAACA	4260
	TAATTGCATC AGCCCCTAAG GCTTTAGCTT GGATTGAAGC TTGGATTGAC TTTCAGTAT	4320
	CATTAGTGCC AGTTCCTGCT ATGACAGGAA CACGTTTATC TACAAGATCA ATAAGTCTT	4380
35	TTAGAATGCG TTCTTTTTCA TCTGTTGTTA AAGTAGGGCT CTCAGCAGTA GTTCCATTAA	4440
	CGATGATTGC TTGGGCATTA TTTTCTAGTA AAAAATTAAC GTGTGTTTTT AAAGCTTCAA	4500
	TATFAACTTT GTTATTGTA AAAGGGGTTG TAAGTGCaAC sCCAACACCC TCAAATAAAT	4560
40	GTGTCATTTT AATTCGCTCC TTTTAAACGC ATAACCTGTT CCAATACTTG TACAGCATTT	4620
	AATGCAGCAC CTTTAAATAA ATTGTCTGAT GTACACCATA CATGGAAAGT ATTTTCTAAT	4680
45	GAATCATCTC TACGTATACG GCCAACAAAC ACTTCATCTT TATTAGTAGA ATTGATTGCC	4740
	ATTGGATATT CATTGTTCTC TGGATTGTCT ACTAAAACAA CGCGGTCATC TTGATCAAAT	4800
	AACGCTTTAA TATCTTCTGC TGTTGTTTCT TTGTCAAGCG TTACATCAAT TTCAACACTA	4860
50	TGACTATCTT GAACAGGCAC ACGTGCGCAT GTTGCTGTTA CTTTAAAGTC TGGCGCATTT	4920
	AAAATTTTTT TCGTCTCATC AATCATTTTT TGTTCTTCTT TTGTATATCC GTTTTCTAAA	4980
55	AACACATCAA TATGCGGTAA CACATTATTA TAAATTGGAT GTGGATATGC TTCTGGTGCT	5040

	TGATATGTTG TATATGCCAC TCGTTTTAAA CCATAAGCAT CTTGCAATAC TTTTAGAGGT	5160
	ACAACAGATT GAATCGTAGA GCAGTTTGGG TTGGCAATGA TACCTCTTGT AAATGTAGGT	5220
5	TCATTGACTT CCGGAACGAT TAAATCAATA TCTTCTGCCA TACGCCATTG ACTTGAATTG	5280
	TCTATAACGA TTGCACCAGC TTTTTCAAAA AGTGGGGCAA AGTGTTCGCT TGTACCGCCA	5340
10	CCAGCACTCA TTAATACATA ATCGAAATGT TCACTTGCAC GAGCATCAGT TAATTCCTGA	5400
	ACTGTATATG TTTTTCCTTG AAATTCAACT TCTTGCCCTG CAGAACGTGC TGATGAAAAAT	5460
	AATACTAATT CATCGAAAGG AATATTTTAA CGATTTAATG TCTCCAACAT TTTTGTACCT	5520
15	ACTAATCCTG TTGCACCAC AACTGCTAAC TTTGTCTATA CTTGTCACTC CATTTTATAA	5580
	TAATTTCCaA TTTTTAGAAT ATTTTAACAA TCATTTTACC ATTAAATGTT AAATGCGTCA	5640
	TATAGTTTTT CTACCGCTTG TTGCCCATTA AAATCATCAA TGACGTATGA AATACTTATT	5700
20	TCAGATGTTG TIGTTTGGTA GAAAGGTATA TTATTTTCAA TTAATGTCAA AAATGCTTTT	5760
	GATGCCACAC CTGACATATC ACGCATGCCT GAGCCAATTA ATGAAATTTT GACATAATGC	5820
	TCATTGATTT TATAAGCTAA TGCTTCATAT TGATTCTTTA ATGTTTCAAG AATCATAGAA	5880
25	ATTTGATGAA AATCACTATC TTTAATCGTG AAGGATAGTT GTAGCCCATC CAAGTTGACG	5940
	ATTTGTGAAA TCATATCAAC ATTTACAGCA CCTTCTTCAA GTTCCGTAAA TAGTTGGGTA	6000
30	AGTAGCTGAT TGTCAGGTAG GGGATAACTA ATTGTTACAT GCATCATATG TTTATCCAAA	6060
	GCCACACCAG TAACTGCTTT TTTCTCTAAT ATTTCTTCAT TTGACATAAT CCATGTTCTT	6120
	TTACGTTTCG ATAAAGTTTT TCCTAAATAT AAAGGGATAT TATAGTTTTT AGCTAATTCA	6180
35	ACACTTCTTG TTTCAAGTAC ACCAGCACCT AAAGCGCTCA TTTCCATCAT TTCTTCATAT	6240
	GAGACGATGT CTAGTCGTTT AGCCTTTGGT AAAAGTCTTG GGTCAGTGGC ATACACACCA	6300
	TCAACGTCGG TATAAATTTT ACAAGGTATT TGATTACTAA CAGCAAGTGC CACAGCGGTC	6360
40	GTATCAGAAC CACCTCTGCC TAAAGTTGTT AATTCCTGAT GTTCATTGAT GCCTTGAAAT	6420
	CCAGCAACTA CTAAATATC GTTTTCTTGA AAGGCTTGTT CAAATGTTTG AGGATTAATT	6480
45	TGAGCAATTT TACTTTTTAA ATGATGGCCA ATGGTTTTAA TACCCGCTTG ATAGCCAGTC	6540
	ATTGCTTTGG CATTCATACC GATATCATTT AATACCATTG ATAAATAAGA TACAGTTTGT	6600
	TGCTCTCCGG TTGTCAATAA TAATGCCAGT TCTTGTGTGT TTGGTGCTTT AGTCAAGGTT	6660
50	GATACATTCG TCATTAATTG ATCTGTTGTG TTACCCATAG CACTTACAAC GACAATTAAa	6720
	TTGTTTCATCT TGATGACTC GCTCCTTTAA CATTTTCAGCG ATCCTTTTTTA TTTTGTAAa	6780
	ATCACTGACG GATGATCCGC CAAATTTCAA CACACTTCTT GTTACCATAT AATCCTCCTA	6840

55

TAATCTATAT ACAAGTGATG CACTCCATTA TTTTAAATA ATGACAACT CTCAGCTCTT 6960  
 AACCAAAAAG TCCAACAAAT TATAACTGCT ATTATAATTG CTTGGGCATC GCACCCCTTC 7020  
 5 AAATTTAGCT GTTAGCAGAC AGTAATCTAa ACTTTACTCA TGATTGATGC GCCTC 7075

(2) INFORMATION FOR SEQ ID NO: 264:

(i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 5171 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

AGACGTACTT TGTGATTmCG AAGyrCGTAC TmAGCACTT GTCGACGTTG ATGTACTTGT 60  
 20 TGAACCTGAT TGAAGTAGTAC TTTGTGATAA TGACTTACTA TCAGAATCAG ATGTACTTTG 120  
 TGAATCACTT AATGATTCTG ATGTACTACC TGACTGAGAC GTGCTCATTG AACTACTTAC 180  
 GGACATTGAT TTAAGTCTG ATGCAGATAA TGACCCACTT GTACTGATAG AGTCACTTAC 240  
 25 TATCTCTGAA GTACTCATCG AGTCTGATGT ACTTGTGAG AACTTTGTG ATGCTGCTAT 300  
 GCTTAGTGAT CCAGAAACAG AACCACTTGT GCTCGTCGAA TCGCTCAATG ATTCTGATGT 360  
 ACTCATCGAT TTTGAATCAC TTGTACTTAA TGATATTGAT GTACTTTGTG AATCTGATTT 420  
 30 GCTTGTGAC GCACCTTGAG AGTTGGCTAT GCTATTTGAA ATACTGATAG AGTCCGAGGT 480  
 GCTAGCTGAC TCGCTCAATG ATGTTGATGT ACTAATTGCA TTCGATGTAC TGCTACTTAA 540  
 35 TGATGCTGAT GTACTAGACG ACCCTGATAT ACTCGTTGAT AAGCTTTGTG ACTTAGACAA 600  
 GCTTCCTGAT GTACTCATAC TTAATGAGTC ACTGAGTGAT GTTGATGTAC GCAATGAATC 660  
 AGATSTACTT GTTGATAGAC TTTCGGATTT TTCAGTACTG CTAGAGTTTG AAATAGAATC 720  
 40 GCTTAATGAT GTTGATTAC TAGCTGAATC CGACATGCTT GATGATACAC TTTGTGAATT 780  
 CACTAAACTT GTGCTTGTG AGCTTGATAC ACTATTACTT TCAGATGTGC TTAATGACTT 840  
 AGATGCACTC ACAGAATCAG ATAGGCTTAC ACTTGTGAT TTCGAGGTAC TAGCTGATGT 900  
 45 AGATACCACA ATCGATCCTG ATGTACTCGT TGATGCACTT TGTGAGTCAG CTTTACTTGT 960  
 TGACACACTT TGAGATTGTT GTGTACTTCC TGATGTTGAT ACGGAATCAC TCATGCTATT 1020  
 50 TCTTGTTACT TCATATTTAA AAGTTGTGCT CGTTTTGTTA CCGCTCGCAT CTGTAGAAAC 1080  
 GATTGATATA GTACTTGAT CAATGTTTGT TGGTGTACCA CTAATAGTAT TATTTGACT 1140  
 ATCAAATGTT AGTCCGATG GCAATCCAGT CACTGTATTC GTCACCGCAT TTCCACTGTT 1200

55

	ATTGGTGTCA CTGTTGGTGC TGTCGTATCC ACAACATTTA TTGTAAAAGT TGTCGTGAT	1320
	TTGTTATTTG CTTGGTCAGT AGACACAAC TACTACTGTTG ATTGACCAAT TTTTGTGGT	1380
5	GTCCCAATGA TTGAATTCGT TGCACTATCG TAACTTAATC CGCTTGGTAA TCCTGTAACT	1440
	GTATTTGTCA CAGTCCAGT ACCATTATCC GTTGTAGTCA ATACAATAGG ATTCATTGTT	1500
10	TTACCCACTT CTATGGTTTG ATTGCCTACA GTTACAGTTG GTGCTTTTAC ATCAGTAAAA	1560
	TAATATGTCA CTGATTGTCC AGCATTCGTC ATTTTACAG TTTTATTTGT ATCATTATAA	1620
	GTTGACGCAT ATGAACATC GACGGACGTG TAGTTATATC CTTTAGCAGT CAATGCAGAT	1680
15	TGCTGaTTAT CGATTGTCAC GACTTGATCA ACATTTCTTG AATATGTTTT TGGTGGAATA	1740
	ATATCTTTAC CTGTTGTTAC ATCAACGTAT CTCACTTGTG TAnCAGCAGA CTCTGTATAT	1800
	TCGAATGTTT CAAATTGTAC TTGTTGTAAA TTTGTCGCGC CACCTGTTGA GGCTGTCATT	1860
20	GATAATGAAA AGTTGGTCGT ACCACTTTTC GCAATCCAAT CTGAAATATT ACGTGTCCAT	1920
	GTTTGACCTG CATATTTGAC AGTCATAACC TTTGTATCAC CATTATAGTT AATATCAAAA	1980
	TCTTGaACG TGTTATTTGT AGGTTGaACA TTTAACTTCG CAGCATTATC AGCTGTTGAA	2040
25	CTTGATGTAT ACGTTGTCGC AACACCATAA CTATCTGTTG TTACAAATGC ACCAAACGCA	2100
	CCTCCACCAG CTACATTAGA TGGGTCAGCA TTCGCCTTTG CAGCTGAATT TGGTTTAGAT	2160
30	GTATTGTGAT ACGTATCCAA TTTGAAGCCA AATGCGTTAC TTAAGCCACC AATACCTACT	2220
	GCGGCACCGT TTAACCCTGT TTCACCTAAT ACACCTGGTG AAAAGGCAAA ACCGATACCA	2280
	TCTCCACCAT TTCCATGCCC TTCATATTTG TTACCTAAAT TTACTTTTCC AGAAAAATGA	2340
35	AAACTCTTAT TAGAGTCAAT ACGTGTTCCT AATGTAATAG CACCTTTTGT GCTGTATGCA	2400
	TCCTGTGTTA ACGTCACAAT ACCGGTACTT TGATCATAGG TAGCATTACC TGACGTTGTC	2460
	ATATATTGTT TTAAGTTATC TTTATTAAC TTAATTGTAT TAGCAGTTAC TGCGGTTGTC	2520
40	GTGgCTGctG ACGCAAATGT TGACATAGCT AAGCGACTGA AAGTTCGAAG TTTTACTGGT	2580
	GCGGTGCTAG TTGACGTTGT GCTAGTTTTG TTTAAGTTGA CCGAAGATGG CGTTGTGCTT	2640
	TGTGAAGTGT TATTTGATGC AGTACTTTGA TTTGTTGATG TATTAATTGG TTGTTCTGTA	2700
45	CTTGAAGTTG AAGCTACAGA TTTAGTATCA GAACTTGATG TAGTATTCTT TGAGGATGTT	2760
	GATTCTGATG TAGATGTCAA TTTCTCTTGT TGATTGCTTG TACTATTAGT TGTCGAAGTG	2820
50	ACCTTTTCAG ACTTTTCACT TGAGACTGTG TCACTATTTG ATGTTTGTAC CGAACTACTA	2880
	TTTtTCGTTA CACTTGTTGA ATCGGCTGTT GATGTTGATG CTTGATTGT CGTTGAGTTT	2940
55	TGATTACCTA CTGTTTCACT TTGTGTGTTT AATTCAGAAG TTAATGGTGC ATCAGAAGCC	3000

	GTTTTCAGTC CGTATCCCGT CATTTTTTTTA CTAATGCTTT GATTATCTTG ACTCACTAAA	3120
	CTATGACTAA TAAATGGTAG CCCCATAAAT TTGAACATTT CTATTCTTT AATTCCGGAT	3180
5	TTTACCCAAT TTTTTCAGA TTTATAAAGT CTTACTCTTG TTTTTCGTT TGCTAAGCTG	3240
	TCATGAAATG CTTTCTGTCT TTTACTCATG TAATAACTCC TTGTATTATC TTTACATTCA	3300
10	TTAGATTATA ATATATGCCA CTATTCAATT TAATACAAC TTTTTTTGAT ACAAAAATAC	3360
	TCATTTTGTT AAAATTTGTA AAAATTCaTT TTTATTCGTC TAAATGTAAT CGTTTTCATA	3420
	TTTTTAAAAT TACTTTTTCT CGTTTATGCG TATAATCTTT TTTTATATAA ATTTGGCTAA	3480
15	TTGGCTTTAT GTTTAATCAT TATAATTGTT TCGTTTTTAA AATAATTATT GTATTAATAT	3540
	ATCTATACCA TCCACCTTTT ATTTATAAAT AGTTAAITTA CAACTAAACG ATAAATATTA	3600
	TATGCAAAAT ACATCTTTAA TATTAAAGTA ATACCAATAT TTTTCAATA AACCTAGTGT	3660
20	AATATATGTG TAATTCTAAA AGATTCTTCT TTAAAAATAT AAATACCACG ACATATTGCT	3720
	TTAACATTTT CATTTATAAA GCGAAAAAAT GCATCGCTAC TAAGTTGAAT GTTTAGTAAG	3780
	GATGCATTGA ATTCACTAAA ATGATTAAAT TACTTATATC TTTTCATCTG ATTGATTATC	3840
25	GAAATTTCTT CCTTCTAAAC CTGCTAACTC TTCTTTAGAA GCTGCAGGTG CTTTCATTTC	3900
	AAATATCTCA TTCACTACTG TGTAATCGTA ATATCTAAT CTGGCAATAG GTTTAATCGA	3960
30	CTTAATGTCC AATTTACCAT TATCAAGAAT AACCTTATCG TCAATATGAA CTTGGGCAAC	4020
	TCTTCCTATA ACAATATCTA CGGTAGATAC TGGATCTCCA GTTGAATAC GAATCGTTTG	4080
	AACGTACTCA CATTCAAAAT GAACTGGCGA TTCTTTTACA CGATATCCTG GAGCTTCTAT	4140
35	ACATTTTTC TTTGTTACAC CTGCAAAATT AAATTCATCC TCTTCTGGTG GCAATGCTTT	4200
	CGATGATAAA TTAAGTCTT CTCTTAAATC ATACGTTGCC ATATTCCACA CAAACCAACC	4260
	TGTCTCTTCA GCATTTTTC CTGTATCTTT ACGTTCGTGA TCACCAAGAA CGGATTGATT	4320
40	TGCTGCGAAC ATAACCATAG GCGGATCCCA AGTTAAGTTT TGATACTGAC TATAAGGCGC	4380
	TAAATTATCT TTCCCATCTT TCGATACAGT AGAGATCCAC CCTATTGGAC GTGGTACTGT	4440
	ACTACTTTTA AATGGGTCGT GCGGTAAACC ATGACTTCTT ACACCTTGTT TTGGCGAATA	4500
45	ATTCATACTA TCTTCACCCC TTATAAGTAA TTACATTTAA GGTTACGCCC TCTTTACATA	4560
	AGCGTCTAAT ATAAATAAAC AATTATTTTA TAAGTAGAAA CTATATATGA CGTGGTTGCT	4620
50	TATAATTTGC GTTCTTGATT CGAAAAATTC AGATAAGGAT TTATACAATT AATATTTATG	4680
	ATATCTTTTG TAAATTTAAT TAATTATAGT TACTTCAATC ATGATTAGTT TATAATAATA	4740
	AAGTGAAATT GAAAAAGACA GCTATTATGC GATGAGCGAA AACTTCAAG TAAAAACAAGA	4800
55		

TAAGTTCAAA AAAGAATTCA AACCTGTTAT GCACTTAAAA GGTGATGCAT TCAATCAACA 4920  
 GTTACAATCT TTGATTAAACA AATATCCACA AATACAAAAA AATATGAAAT CAGAGTTCAT 4980  
 5 TGCTTATTAT GATAAAGAAA AAAATAGAGA AACAGTAAAA AACTATGCTT GGAACCTTCA 5040  
 AAAATCTATA AATGACATTA TGCAATCATA TCCTAGCACA AAATTGTAC AGTTTTATAA 5100  
 AAGATGATGT TTCCCCGTCA ATGGTAGATG GAAATGGCCG TTTAAATCG GGATACTAAT 5160  
 10 GTATTTCCAT C 5171

## (2) INFORMATION FOR SEQ ID NO: 265:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3589 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

CTACACACTA AACCTATTTT AGTTATGGGT GGTACAGTGA TTCTCTTTTC ATTTTAAATA 60  
 25 GGTATTTGGA TTGGTCATCC TATTGAAACA GAAATCAAAC CACTTATTAT TGGTGCATG 120  
 ATTATGTACG TACTTGGGCT TGTAGATGAT ATCTACGATT TGAAACCGTA TATAAAATTG 180  
 GCTGGTCAAA TGCCGCTGC CTTAGTAGTT GCTTTTTATG GTGTGACTAT TGATTTTATT 240  
 30 TCGTTGCCAA TGGGTACAAC GATTCATTTT GGATTTCTTA GTATTCCAAT TACTGTGATT 300  
 TGGATTGTTG CTATTACAAA TGCAATTAAC TTAATTGATG GACTCGATGG TTTGGCGTCG 360  
 GGTGTTTCKG CAATCGGACT CATTACAATA GGGTTCATTG CAATTTTACA AGCTAATATT 420  
 35 TTCATAACGA TGATTTGTTG TGTTTTATTA GGCTCTTTAA TTGGGTTTTT ATTTTACAAT 480  
 TTCCATCCTG CCAAAATATT TTAGGTGAT AGTGGGGCTT TAATGATTGG ATTTATCATC 540  
 40 GGATTCCTTT CTTTACTCGG ATTCAAAAAT ATTACAATTA TTGCATTGTT CTTCCCAATT 600  
 GTTATCTTAG CAGTTCATT CATTGATACT TTGTTGCGAA TGATTGACG TGTGAAAAAA 660  
 GGGCAGCATA TAATGCAAGC TGATAAATCG CATTTGCATC ATAACTATT AGCTTTAGGC 720  
 45 TACACACATA GACAAACAGT ATTATTAATC TATTCAATCT CTATTTTATT TAGTCTTTTCG 780  
 AGCATTATTT TGTATGTATC GCCACCATTA GGTGTTGTAT TAATGTTTGT ATTAATCATA 840  
 TTTAGTATTG AATTAATTGT TGAATTTACA GGATTAATAG ATAACAATA CCGACCAATA 900  
 50 TTAAATTTAA TTAGTCGTAA GTCATCTCAT AAAGAGGAAT AGGGAATGAA AGCATAGCTG 960  
 TATGGGATAA TTTGTATTAT ATGGCTTTAC TCTTTACAAT TTTTTTGTAT TAAATTTCAA 1020

55



	ATTTACCGTC TTATGATAGT GCTTTTATT TTTATTCAGT TGGTATATCG AAAGGTAAC	1140
	GCTTTGGAGT TTCTTCAGTC AAATCGAAAT TTCCTGCAGT CATTTGATTT AAAAAGTTAA	1200
5	TAAACGCTTC ATAGTCACTT TTAACGACAT CGATATAGTA GCTTACCTTA TCAGTGTAAG	1260
	TTTGGTTTCT TAACATAAAA TGAGTTGAAG CTAATTCATA TTCAAATTTA CCAGTTTGAT	1320
	CATAATTCAG TGTTACTATA CATGGTACTG CTTCTCGTAG TTCGACACGC CCGATATCAT	1380
10	AAATGACGTC TCTAACAGCA CCGCTATAGG CGCGAATTAA ACCGCCACCA CCTAATTTAA	1440
	TACCACCAA ATATCTTGTT ACTACGACAC ACGCATTATG aACATCGrGC TTTTTTaATA	1500
15	TGTCTAACAT TGGGaCACCG GcAGTTCctG TCGGTTcACC ATCATCATTC GcTTTTTGAA	1560
	TATTCATTtC AGGTCCAATA GTATATGCAG AACAATTATG AGTGGCATCT TTATGTTCTT	1620
	TTTTTATTGC AGCAATAAAT GCTTTaGCTT CATCTTCATT TTGAACAGGT TGATATGAG	1680
20	CAATGAATCT TGATTTACTA ATCACATTTT CAATAATGTG TTCTTTTTTA ACAGTAATGA	1740
	TATTTTGTGT CATAATAACT CCTTAATTCA TAAGCTTAAG ATTATTTAAT CTTCATTATA	1800
	CACTGAAAT GACATGACTA TAAATCGTTT GATTGCCATT TTCTTTTTTA CTGAAATATT	1860
25	GTATCATTGC TATGAGTATA TTTTAGGAGG ACGACTATGA AAATTGCTGT GATGACCGAT	1920
	TCTACAAGTT ATCTGTCGCA GGACTTAATC GATAAATATA ATATTcAAAT AGCGCCATTA	1980
30	AGTGTGACTT TTGAAGATGG CAAGATTATA CCAGAAGAAA AAGTTCGTAC TAAAAAGCGT	2040
	GCCATTCAAA CATTAGAAAA GAAAGTATTA GATATTGTAA AAGACTTTGA AGAAGTAACT	2100
	TTATTTGTCA TAAATGGAGA TCATTTcGAA GATGGTCAAG CGTTATACAA AAAGTTACAA	2160
35	GATGATTGTC CTTcAGCTTA TCAAGTAGCA TACTCTGAGT TTGGTCCAGT TGTTCAGCA	2220
	CATTTAGGTT CTGGTGGATT AGGTTTAGGC TATGTTGGCA GAAAAATAAG ATTAACATAA	2280
	TTATAAAATT TTAATAAAAG AGTCTATATT GTAATTGGAA ATTATCTCTC GTATACATGG	2340
40	CTTTAAATGT TCATCATTTG AAAGCCAAAA TGCTAAAGAT ATAAGAAAAT CATTATAATA	2400
	TTAGGCTCTT TTTIACGTTG AAATGAGGTT TTAAGCATTa AACATTACGG GAAATTAATT	2460
	CATCCTCATA CTTCACTTAC TAATGAAAAA ATTAAAAAAG AAGTAACAGG TGTCATCAAA	2520
45	CAAAATTCAA ACTATTATTG TGTTCAATGT GAAAGTACAA ATCCAAAGCA TTTTATCAG	2580
	TATGATTCCT CAGTACATTC CAAGAAAATT GTATATTGCA GAAATTGTAT ATCACTGGGT	2640
	CGAATGGATA ATGTAACAAG ATATAAAATA ACAGAGAGTT CGCAAAGTTC ATCACAAGCA	2700
50	TATTATCATC TCTCATTTGA ATTGTCGGAA CAGCAGTCTT ATGCCTCAGA ACATATTGTT	2760
	CGAGCCATTA GAAAGAGACA AACGATTTTG TTATATGCCG TAACAGGTGC AGGTAAGACA	2820

55

TCACCACGTG TAGATGTTGT TGTAGAAATT AGTAAACGTA TTAAAGACGC ATTTCTTAAT 2940  
 GAAGATATAG ACATACTACA CCAGCAATCA AGACAACAAT TTGAAGGGCA TTTTGTGTGA 3000  
 5 TGCACAGTGC ATCAACTTTA CCGATTCAAA CAGCACTTTG ATACTATTTT TATTGATGAA 3060  
 GTCGATGCCT TTCCTTTATC AATGGATAAA AATTTACAAC AAGCATTGAA GTCATCTTCT 3120  
 AAAGTTGAAC ATGCAACAAT TTATATGACA GCAACACCAC CGAAACAAC TCTGTCAGAG 3180  
 10 ATTCCCCACG AAAATATAAT TAAATTGCCA GCTCGCTTTC ATAAAAAATC ACTTCCAGTT 3240  
 CCTAAATATC GTTATTTCAA ACTTAATAAT AAGAAGATTC AGAAAATGTT ATACCGAATT 3300  
 TTACAAGATC AAATTAATAA TCAACGTTAT ACACTGGTGT TTTTAAACAA TATAGAAACA 3360  
 15 ATGATTAAAA CATTTTCGGT TTATAAGCAG AAAATTACTA AATTAACATA CGTCCATAGC 3420  
 GAGGATGTTT TTCGCTTTGA AAAAGTTGAA CAATTAAGGA ATGGACATT CGATGTCATT 3480  
 20 TTTACTACGA CAATATTAGA ACGTGGATTT ACAATGGCAA ATTTGGATGT TGTGTGTTATC 3540  
 GATGCACATC AATATACTCA AGAGGCTTTA ATACAAATTG CTGGACGTG 3589

## (2) INFORMATION FOR SEQ ID NO: 266:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1017 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

TTTCCAAGAA GGcTTgAAAA AtGTTwCaAG TGGTGCgAmC CCAGTTGGTT TACGACAAGG 60  
 35 TATCGACAAA GCAGTTAAAG TTGCTGTTGA AGCGTTACAT GAAAATTCTC AAAAAGTTGA 120  
 AAATAAAAAT GAAATTGCGC AAGTAGGTGC GATTTGAGCA GCAGATGAAG AAATTGGACG 180  
 40 TTATATTTCT GAAGCTATGG AAAAAGTAGG TAACGATGGT GTCATTACAA TTGAAGAATC 240  
 AAATGGACTA AACACTGAAC TAGAAGTGGT TGAAGGTATG CAATTGATC GTGGTTATCA 300  
 ATCACCGTAT ATGGTTACTG ATTCAGATAA AATGGTTGCT GAATTAGAAC GCCCATACAT 360  
 45 TTTAGTAACA GATAAGAAAA TCTCGTCTTT CCAAGATATC TTACCTTTAT TAGAACAAGT 420  
 GGTCAATCT AATCGTCCAA TCTTAATTGT AGCTGATGAA GTTGAAGGCG ATGCATTAAC 480  
 AAATATCGTG CTAAACCGTA TCGGTGGCAC ATTTACAGCT GTTGCAGTAA AAGCACCTGG 540  
 50 TTTTGGTGAT CGTAGAAAAG CGATGCTTGA AGATTTAGCT ATTTTAACTG GTGCGCAAGT 600  
 GATTACTGAT GATTTAGGCT TAGATTTAAA AGATGCATCA ATTGATATGT TAGGTACTGC 660

55

CAGCATTGAT GCACGTGTTA GCCAATTGAA ATCTCAAATT GAAGAACTG AATCTGACTT 780  
 TGATCGTGAA AAATTACAAG AGCGCTTAGC TAAATTAGCA GGTGGTGTG CAGTTATCAA 840  
 5 AgTAgtGTGCA GCAAGTGAAA CAGAGCTTAA AGAACGTAAA TTACGTATTG AAGATGCATT 900  
 AAATTCTACA CGTGCAGCAG TTGAAGAAGG TATTGTTGCA GGTGGTGGTA CTGCATTAGT 960  
 10 AAATGTTTAC CAAAAAGTAA GTGAAATTGA AGCTGAAGGT GACATTGAAA CAGGTGT 1017

## (2) INFORMATION FOR SEQ ID NO: 267:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1409 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

TTAATCCAGC GTTAACTGTA TTTGCATTTA TTATGATTAT TTCGATTCTT TTAGCGTATG 60  
 TATTTAAATG GCTTGGATTA GTGGATGATG TGTATTAAAT GGTCATTATC ATTTCAACTA 120  
 25 TTTCCTTAGG CGTAGTTGTT CCAACTTTAA AAGAAATGAA TATTATGAGA ACAACTATAG 180  
 GGCAATTTAT CCTATTAGTA GCAGTACTTG CGGACTTAGT AACTATGATT TTATTAACGG 240  
 TCTATGGCGC AATCAATGGT CAAGGCGGCA GTACAATATG GTTAATAGGT ATATTAGTTG 300  
 30 TTTTCACAGC AATTTTCATAT ATTTTAGGTG TTCAATTTAA AAGAATGTCA TTTTACAAA 360  
 AATTGATGGA TGGTACGACG CAAATCGGTA TTCGTGCGGT ATTTGCATTA ATAATATTAT 420  
 TAGTAGCCCT AGCAGAGGGA GTTGGCGCAG AAAATATATT AGGTGCATTC TTAGCAGGTG 480  
 35 TCGTTGTTTC ATTATTAAAT CCAGATGAAG AAATGGTTGA AAAGTTAGAC TCATTTGGTT 540  
 ATGCGTTCTT TATTCCTATT TTCTTTATAA TGGnTGGTGT AGATTTAAAC ATACCTTCAT 600  
 40 TAATTAAAGA ACCGAAATTA CTAATTATCA TACCGATTTT AATCGTnGCA TTTATCATTT 660  
 CAAAATTAAT TCCAGTCATG TTTATTGCGC GTTGGTTTGA TATGAAAACA ACGATTGCAT 720  
 CAGCATTTTT ATTAACATCA ACATTATCGC TCGTGATAGC TGCAGCCAAA ATTTAGAAA 780  
 45 GATTAAATGC TATTTTCAGCT GAAACGTCAG GTATATTAAAT TTTAAGCGCA GTCATTACAT 840  
 GTGTATTCGT TCCGATTATT TTCAAAAAC TGTTTCCAGT TCCAGATGAG TTTAACCGTA 900  
 AAATTGAAGT TAGTTTAATT GGTAAAAATC AATTAACGAT TCCTATAGCG CAAAATTTAA 960  
 50 CATCTCAGTT ATATGACGTG ACATTATATT ATCGCAAAGA CTTGAGTGAT CGTCGTCAAT 1020  
 TGTCAGATGA TATCACGATG ATAGAAATTG CTGATTATGA ACAAGATGTT TTAGAACGAC 1080

AAGTTGCTAA ATTAGCCAAA GCACATCAAG TTGAGCGTGT CATTTGCAGA CTTGAAAGCA 1200  
 CAACGGACGA TACAGAGTTA GTTGATTGAG GTATTGAAAT TTTCAGTAGC TACTTAAGTA 1260  
 5 ATAAAACTTT ATTAAAAGGT TTAATTGAAA CACCTAACAT GTTGAATTTA TTAAGTAATG 1320  
 TTGAAACGTC ACTATATGAA ATTCAAATGT TAAATTATAA ATATGAAAAT ATTCAATTAC 1380  
 GTAATTTCCC ATTCGGAGGA GACATCATC 1409

(2) INFORMATION FOR SEQ ID NO: 268:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4702 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

AAAGAGGGTT TTTTACTACT ATAATCCATC CTTTAATGGA ATTCCATTG TCCCTTTGGC 60  
 CCGTTCATAA CCATAGGAAA ATATATGCGG AATCATTGA TAGTAAATAC ATCAATAATA 120  
 25 CGAATAATGT CCGTGAAATC ACGTCCGTAT TACCATTATT AGCTAAAACA TTTTCCAAGT 180  
 TTTCTTTTGA ACCGGATACT CATAAATGCT TTTAATGCGT GGTTTTGTCT GTCGCCATCT 240  
 TTAGAATTG TAATACTTTT TTAAATTCTT TAGCGAACAA CTCATTATCT TTATCGTTTT 300  
 30 TAGCCATTG ACGATTCAAA TCGTTAGCTC TTACGCCTAA TGCTCGACCT GTCGCATGTA 360  
 GCTTATTCAC ATTATTGTCG ACTTGATTGA ATTGGCCACT AACAGAATCT GCAATTGATT 420  
 TTGATTCTTG TGTATCTGAT AGCAATTGCG TACTCTTTTC AGAAATTCTA CTAATTTCTT 480  
 35 TATCTAAATT TGAAGACATC GTATTAAATT CATCATTTTT GCCTTTATCA ATTTTGGTT 540  
 CTTGTGGCTC TTCAGCAAAA GTCTTTTAA CGTTTCTAA CTGATCAATC AGCTTGGAAA 600  
 40 TATCTTCTTt ATTTTGGTT gTATTCTTTT TGTTATTAA AATGTCATCA ATCAGTTTGT 660  
 CTGAGTTTTT TTCCATTGAA TCAATTTGAT GTAACACAGC TACTTTATCG TCTTTGAAAC 720  
 TTTCCATGTC ATTGATAACT TGGTCAACCA TCATATCAAT TAAACGTTT TGTGCAATG 780  
 45 GTTTATCTTC tCTGCCTTTT GTATCTGTGT ACATTTTATA ATGCGCATCA AACCTAGATA 840  
 ATGCACTCAA TTGCTGGCTT AATGCATCTT TCGATAAGCG ACCATCaAGG TTATGATTCA 900  
 ATGTTACATC CACAACACTC GTTGCTTTCT TATCATTGG TTCATCTTGA CGATTGCTT 960  
 50 GTCCAAATAA CAATTGTAAA TGCATTGTTT TATCTTTTAA GAAATCTTTC TCAGCATCCT 1020  
 TTTTCAATTT AGCAACGCCA TTGACTTCAA CTTTATATTC CTTGTTAGAT GTATCGAGTT 1080

	TAAAAATGAGG	AtCTGTTGCA	ACAGTTAATT	GATTAAATATC	ATTACTTTTA	ATCGTTTCAG	1200
	TACGTTGCAC	TTTGACACCA	TCATTAAATCA	AACTACTTGT	GTCTTGCGCA	ACTTGGTTAT	1260
5	CGTAATCTGT	TAAATTAATG	TGTTCTGCTA	ACGGTTTTTT	CAAATTATAT	TCATTTTTAT	1320
	AACGTTTTGC	TTCTTTGACA	ATTGCTTCGT	ATTTATTAGC	TTCATCCTCA	TTTAAACCTG	1380
	CAGCTATAAA	GTCTTGTTTA	GACATGTTAT	AGATAAATGT	TGTATCTGTA	TCAGGTTCTT	1440
10	TGACAATATC	ATCATGAAGT	TGTTTCTCTA	AGTTTTCAGC	GAATTGAGCA	TTGTTCAATT	1500
	TAATGCTATT	TAGCGCATCT	TGTAAGTCTT	TGTTATTGTC	AAGCTCATCT	TGCAGTGATT	1560
	CTGTTAATTG	CTTACGATA	TCTTCAATCA	TACCTTTTGA	AAATGGTGAC	TCTTGTGATT	1620
15	GAATGATTTT	TCTTAATTTA	TCTAAGTTTT	CTTTAACAGT	TTGTTTATAT	TCTTCTTTAC	1680
	CTGTATCTTG	CATACTTGAT	TGTTGATCAA	TTTGGCTGTC	CATCTGTTTT	AATGCATTGA	1740
20	TATAGTTATC	AAGTTCCACG	CTATCTTTTT	GCGATTTATA	ATCTTGTAAC	ATTTTATCCA	1800
	TCGCTGTATT	GTGCTCGTCA	AATAATGAAT	TTTGTTTTTT	AATTAAAGTC	GAAACATTAT	1860
	AATCTGTGTT	CACTCTGAAC	GATCTGAAAT	TCGCACTCAA	TAATGATTTA	TTGTATGTTT	1920
25	GGAAACCATT	TGTAATGTCT	TTGTTTGACG	AAATTGAATT	TACAAGCGTA	TCTGTAAATA	1980
	ATTCGGGAA	GTGTTAATT	GGATTTAATA	AGTAATTCGA	GAATTTACTA	TTCACACCAT	2040
	GTTACGCGT	CATAATAGCG	CCAACATTTT	TTTGTGCATT	ATGTAAATTA	TCAATGATGC	2100
30	TTGTAAATA	AATTTGACT	AAGTTTTTGT	TAAAGTCGTT	AAGTACATTA	CTTACAACCT	2160
	TTTCTGTGTT	TTAGCTACT	TCTTCTTTTT	GTCCTACAGC	TGTTTTATAC	TGTAGCGATA	2220
	TTTTCGATGG	TGTTTTAGCG	TCTAATTGCA	TTGCCAATTT	TGAAAAGTTT	TCTGGGATAA	2280
35	CAATCATGAC	TTGGTATCCA	CCATTTTTCA	AACCAGACTC	AGCAACGTTT	CTTGTTACTG	2340
	TTTCAAAATTT	ATAGTTTTTC	TCATTTGCTA	ACCTTTTAAT	AAATGCTTGA	CCCAGCTCAA	2400
40	CTTTTTTACC	GTTATATGTC	GTTGGTTGAT	CCTCGTTAAC	AATTGCGATA	TGTATTTTAT	2460
	TATTTTTATT	ACTTACACTT	TGGGATCCTT	TTTCTGATTG	ATCTCCATAT	TTTGTTTGAA	2520
	CAAAAAATAT	CATACTAACT	ATGGCAATTA	TAATAATTAA	AGTGACAATT	AATGCATAAA	2580
45	TCCAATTTTT	CTTTTCATG	CTTATTTCTT	TTTCAAGTTT	TTCTTAAAAA	AATGATAAGC	2640
	AAAGCCACAT	TAGAAAATGT	GACTTTGCCA	ATTTTCAAGT	GCTTATTGCA	AACCGAAATT	2700
	ATTAGAAAGT	TGTTGGTCTT	GTCTTTGAAC	GGCATCAGCA	GTGCTATTCA	ATTGTTGTTT	2760
50	AATTTCTTCT	AATAATTGTG	CAAATTTTTT	TACTTTAGGA	CTAAGTTGTT	GGAATTGCTC	2820
	TTCGAAACGG	CTGAAAGCTT	GACCTTCCCA	GTTGCTGCA	ATTTTACCTT	GTGCACGkGT	2880

55

EP 0 786 519 A2

	TCTGATTTC	TCTGGACTCA	TCTTAATCAT	TGCCATAACT	AGAAACCTCC	TGAATATTTT	3000
	AAGTTTATCa	AAACTTTT	GGGACACTAT	TTTTTGAAAA	AGTGCTCCTT	ACTCAAATAA	3060
5	TATATAAATT	ATTAGTATAT	GTATATAGTt	TTTTAAGTAT	TTTTAGCTTT	TTTAAAATAA	3120
	ATATATTGAA	TATAACCATA	TATTTTTAAT	TAACCATTCA	TTTTTGTAAT	ATAAATGTGT	3180
	ATACTAAAAT	TAAATTAAAT	ACATAAAGGA	TTAAATGGTT	ATTATGAAGA	AAACAATTTT	3240
10	ACTGACGATG	ACAACTCTTA	CTTTATTTAG	TATGTCGCCT	AACTCGGCTC	AAGCATATAC	3300
	GAATGATAGC	AAAACATTAG	AAGAAGCAAA	GAAAGCACAC	CCAAACGCAC	AGTTCAAAGT	3360
15	GAATAAAGAC	ACCGGCGCGT	ATACTTATAC	ATATGACAAA	AACAACACGC	CAAACAACAA	3420
	TCATCAAAAC	CAGTCACGTA	CAAACGACAA	TCATCAACAC	GCAAATCAAC	GTGATCTTAA	3480
	CAACAATCAG	TACCATTCTT	CATTAAAGTG	TCAGTATACG	CACATTAATG	ACGCAATTGA	3540
20	TTCACACACA	CCGCCTCAAA	CGTCACCAAG	CAATCCTTTG	ACACCAGCAA	TACCGAATGT	3600
	CGAAGACAAT	GACGATGAAT	TAAATAACGC	TTTTTCAAAA	GATAACAAAG	GGCTTATTAC	3660
	AGGCATCGAT	TTAGACGAAT	TGTATGACGA	ATTACAAATC	GCCGAATTTA	ATGACAAAGC	3720
25	AAAGACCGCT	GACGGTAAAC	CTTTAGCATT	AGGTAACGGT	AAAATCATTG	ATCAGCCTCT	3780
	TATCACAAGT	AAGAACAAC	TATATACTGC	TGGACAATGT	ACATGGTATG	TCTTTGATAA	3840
	ACGTGCCAAA	GATGGACACA	CGATTAGTAC	ATTTTGGGGA	GATGCTAAAA	ACTGGGCAGG	3900
30	CCAAGCTTCA	AGCAATGGCT	TCAAAGTAGA	TAGACACCCA	ACACGAGGAT	CAATTTTACA	3960
	AACAGTAAAT	GGTCCATTG	GTCATGTAGC	CTACGTTGrA	AAAGTTAATA	TTGATGGAAG	4020
	TATTCTAATT	TCAGAAATGA	ACTGGATTGG	TGAATATATC	GTTTCATCAA	GAACCATCTC	4080
35	TGCTTCAGAA	GTTTCATCAT	ATAATTACAT	CCATTAAATT	AATCATGACA	TCAATAAAAA	4140
	GCGAECAGTT	CGCAGTTTAC	AATTCGTAAC	ACTGCAAAAT	TGGTCGCTTT	ATTTTGTATG	4200
40	TTATTTCGATT	ATAAAATTAC	AAAGAAATGT	TCTCTACATT	CCCCATTAAT	CAAAATCGTT	4260
	TACGAAAGTA	TAATTGTAGC	TATAATAATC	CAAGTCGTAA	CAACTAGTGG	CACTATCGTC	4320
	TTGAATAAGA	ATATACCGTA	TTTTTCTTGG	CGATATATAT	CCAGTACTAG	CCAAATTAAA	4380
45	ATGATTATAA	CACCAACAAA	AATAAATACA	GGATTCATCG	ATATAGCATC	TGCCTGTAAC	4440
	TCAGGTTGCA	TTCTTAATTT	AGTGATAATT	AACATCACTA	CTGAAATAAT	GAAAAAGTAG	4500
	ATACCTCTTA	TCTTTGATGT	CTGTAAATCT	AATTCCTGCT	CTTCAATGAC	CTCTTTAGAT	4560
50	TCACCCAATT	CTTTTGCAAT	CAAATAATTT	ACTACCTTAG	GTTTCACCCA	TAAACACTTA	4620
	ATTGCAAAGT	ACATAAAAAT	ATATGATCCA	GTATCCATAA	ACATTAAAAA	GTTGCTTAAA	4680

55

## (2) INFORMATION FOR SEQ ID NO: 269:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2004 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

	ACAAAAAATT CACCCTCATT AATATTGAAA CTAATGTTAT CGACAGCAAC ATGTTTGCCA	60
15	TAACGCTTAG TTACATTTTT AAACCTAATC ACTTTGCCAC CTCTTTTTTT CTCATAGCAT	120
	AAAACCGAGA TTATATGTAT GTATTCCCTA TTTAACCACG TTTATTACAA TTTTCAAATT	180
	TAAATGATTT ATCCTTGAAC TTTTTTAATA AAATAATGAA TAAWAGGWAA TCWCCAGTTA	240
20	AGAAATAGTG TTATTTTACC TTGAATTCAA AAAACACCCC AGTAAACAA GGAATGCTTA	300
	CTAGGTGTCT TCACTATACT TTGGCTTTAT AATTTTGAAT CGTTCTAAA AATGCTGGAC	360
	AATAATGTTT TAATTTGTAA CTACCTACGC CATCMATATT AATCATATCT TGTTCGAAG	420
25	CAGGCTTACG TTTAGCAAAT TCCTCCAACG TGTAATCAGA AAATATACTT ACAGGTGCTA	480
	TCGTAAATTT GTCACCTAAC TTTTACGAA CTTCTACCAA CTGACTGAAT AATACTCGGT	540
	CAACCCCTTC AACCGTATTT ATAAATACTT TTTCACTCGC TTTTGTCTTA AATGGTGTTG	600
30	TGAATACTTC TACTTCATTA CTGAGTAATT TTTTAATTGA AGTATCACAC ATTAATATTT	660
	CGTCATTTTC ATTAAAGAAC CCTTTGAATC TTAATTCATC TATTAAGTGA CTTAATTCTG	720
	ATGTTGTGTA ACCTTTCATT AAACCATGGG TTGAAATTG GTCATAACCT TTATACTTAA	780
35	TATAATCTGk TGACTCTCCT CTTAACACTT GAATGATAAC ACTATAACTC TCTTGTTGTT	840
	TCATACGAGC GATGCAACTA ATAATCATCT TAGCTTCTTG TGTCATATTA TATGATTTAT	900
40	CTTGTTGAAC ACAATTACTA CATTGTTTAC ATTCTTCTAA TTTTTCATTC GGTTCAAAAT	960
	AATGGACAAT TGTGCTTCT AGACATTTTT TTGTTTTTGT ATATTGAATC ATTTTAGTTA	1020
	ACTTTTCGCC CATTTTATCT TTATAGTCAT CATCAGCTTG AGAGACTGTT ATAAAATACT	1080
45	CGTGTAATTT GATAACGCGT TCGCTAAATA ACAAATACA TTCACTTTTC AACCCGTCAC	1140
	GACCTGCACG ACCCGCTTCT TGATAATAAG ATTCTAAATC TCCAGGCATA TTATAATGAA	1200
	TAACAAAGCG TACATTGGAT TTATCAATAC CCATACCAA AGCATTGTGA GCAACGACTA	1260
50	CTTTAACACG ATCAAATAAG AAATCATTCT GCGCTTCTTC TCTTTCTTTA TTGCTCAAAC	1320
	CTGCATGATA TATAACACTT TCAATTTTCT GACTTTCTAA GGCTTCTTGA AGCTCTTCAA	1380

CCAATATAAA TTTTGTACGT TGATAAGTAG GATTTACTTT AAAAATTAAG TTTCTACGCT 1500  
 TAGTACTCGT TTTAATTTGA TCAGTTTGAG CGATATTTAA CTTTTCTCTA ATATCTTGCT 1560  
 5 GTACTTCAAC CGTGGCAGTT GCTGTCAACG CTATTATTGT AAAATCTTGA GGTAACGTAA 1620  
 ATACTTTTGA AATAACATTT TGGTAACTCG GCCTGAAATC ATGACCCCAT TTAGAAATAC 1680  
 AATGCGCTTC ATCAAACGCG ACTAAGTGAA TCTTTATACG CTGAAGCATA TTTAAAAAAT 1740  
 10 ATCGGTTTTT AAATCGTTCT GGTGCAACAT ACAAAAATTG AATTTCTCCA TTTGATAATG 1800  
 CTTTTTCAAT ACGTTGTTGC TCTTTTGTAG TCAAACACT ATTTAAAAAA GCAGCTTGAA 1860  
 TTCCCATCGC TTTTAATTGA TCCACTTGAT CTTTCATTAA TGATATTAGT GGACTTATTA 1920  
 15 CAATTGTTGT ACCACCTAAC mATAAACCTG GACTTTGrTA GcmTATAGAC yTACCTCCAC 1980  
 CAGTtGGkAA GrCACCAAGC ACAT 2004

20 (2) INFORMATION FOR SEQ ID NO: 270:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2244 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

30 AAAGATTGCT TGCCTTGAGG GTTATATAT CTGACTCAAT TGCCACATTT TTATCAAGAG 60  
 TAGTTGATAA TACTCATCAT AATTATAGCT AATATTATAT TTTTTTAAAA GATAGTGTAT 120  
 GATTTTCTGG TGTTTGTTGT ATACGTCATT AAATTTCAAG TAGTCATTCT CCAAGTTATA 180  
 35 CGTATAACAA ATATTTCCGG ATAAAGTTAG AATAAAATAT TTAGAAAAAT CATTCAATTTG 240  
 CGTAATCGCT AAATTAAGTG TTAAATATAA GACATAAGTA ATTAATTTAA TGTGATATGA 300  
 TGTATTATTA CTTTGCTAAA TAGTAGATAG AACAAAATTT GTAATCGGGA GGTAACAATG 360  
 GATTACGCAC ATTTAAATTT AGAACATTTT TTTGCACGAA ACGACGATTT AGATGTTATA 420  
 AGAGATCGCG CTGATTTTCG TATGATAAAT AACTTCACTA ATGAAATGAT GTATCGTGAT 480  
 45 GGTCAAATTG AAGGCACGAT TGATTTAAAT CAGTACTATT ATAAAAATAG ATCAAATGCA 540  
 GCAAGTTTTA TTATGATGGA TTATAAAAAA GAAACTAAGT AAACGAACAA AAGAATTTTT 600  
 TGTTTTTTAA TACGTGAATA ATAAGATTAT TGATATAAAG GTTTTCAAAG GTTATACAAA 660  
 50 AAGATAAAC ATTTATGATT CGTAGATCAA CGTAAAGTAA TGTTGATAAA TGGTTTAAAA 720  
 CGTTTCATTT ACATTACTGT TTATTTATGA ATATGTAACA ATGCATAGAT AAAATTGTTA 780

55



ACCTAAGAGG TGTGGATATG AATAAACACA AGAAAGGTTC TATTTTTGGA ATAATAGGAC 900  
 TTGTTGTCAT ATTTGCTGTT GTCyCaTTTT TATTTTTCTC AATGATATCC GATCAGATAT 960  
 5 TTTTCAAACA TGTTAAATCC GACATTAAGA TTGAAAAGTT AAATGTTACA TTAAACGATG 1020  
 CAGCAAAGAA ACAAATAAAT AATTATACGA GTCAACAGGT ATCAAATAAA AAGAATGATG 1080  
 CATGGAGAGA TGCATCTGCA ACTGAAATTA AAAGTGCAAT GGATAGCGGT ACTTTTATCG 1140  
 10 ATAATGAAAA GCAAAAATAT CAATTTTITAG ATTTATCAAA GTATCAAGGG ATTGATAAAA 1200  
 ATAGAATTAA ACGTATGTTA GTAGATAGAC CAACGTTATT GAAACATACG GATGATTTCT 1260  
 TAAAAGCTGC TAAAGATAAG CACGTTAACG AAGTTTATTT AATTTCACAT GCATTATTAG 1320  
 15 AAAGTGGCGC AGTTAAAGT GAATTAGCTA ATGGAGTCGA AATTGATGGC AAAAAGTACT 1380  
 ACAATTTCTA TGGAGTAGGA GCCCTTGATA AAGACCCAAT TAAACAGGT GCAGAATATG 1440  
 20 CTAAAAAGCA TGGTTGGGAT ACACCTGAAA AAGCTATTTT AGGCGGTGCT GATTTTCATTC 1500  
 ATAAGCACTT CTTATCAAGC ACAGATCAAA ATACATTGTA TAGTATGAGA TGGAAATCCAA 1560  
 AAAATCCAGG AGAACATCAA TATGCTACAG ATATTAAGTG GGCAGAAAGT AATGCAACAA 1620  
 25 TTATCGCTGA CTTTTATAAG AACATGAAGA CTGAAGGAAA ATACTTCAAA TACTTTGTGT 1680  
 ATAAAGATGA CAGTAAACAT TTGAATAAGT AATTTGATAA GCTACGAGTT GTTTTTATGA 1740  
 CTCGGACATA CTAAAAAGAC GCTTTCTATC TTGTTTTGAT AGAAAGCGTC TTTTTCATT 1800  
 30 AGAGAAAACA CATTGATkGA TAAtCCCaCC aATGCAAgTG GGGcAGGACa TCGATAAAGA 1860  
 ATTACTTTTT CTTTAGAAAT TAGTATTTCT TATGCATGAG TTTTACTCAT GTATTCCTAT 1920  
 TTTTAAGTAC ACATTAGTTA TAGCTAATGA TAAAGAACCA CTACATAATA AATCATTAGT 1980  
 35 GTTTTTTTAT CATTCTGTC CCaCTCTCAT CGTGATTGA AATTTTCAAT TCGGATTTTA 2040  
 ATTTCATCTC TTACACGTTG GAACTCTGAC CAAGGCTTGC CTGCAGGATC ATCAAATCCC 2100  
 40 CAATGTTCTT TCTTAACATT TGTTGGTAAA GAAGGGCAAT TTACGTCTGC ATCACTACAT 2160  
 AATGTAACAA CTAAATTGA ATTTnTAATA ATATTATTAT CGGATTAAAA TCTGATGGAT 2220  
 GATTGATAT ATCAATGCCT ACTT 2244

45 (2) INFORMATION FOR SEQ ID NO: 271:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1371 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

55

ATAAGCAATT TAATTTTGAG TCTACAATGG AGGAATTATC ATCTTTATCA GAGACTTGCC 60  
 AACTTGAAGT GTTGGGTCAA ATTACTCAAA ACAGAGATCG TGTAGATCGC AAATATTATG 120  
 5 TTGGTAAAGG TAAAATTGAA GAAATTCAAG CATTTATTGA GTTCAAAGAT ATTGATGTAG 180  
 TCATCACAAA TGATGAATTA ACGACTGCAC AATCCAAATC ACTAAATGAA GCTTTAGGTG 240  
 TAAAAATTAT TGATAGAACT CAGTTGATTC TTGAAATATT TGCATTAAGA GCAAGAAGTA 300  
 10 AAGAAGGTAA ATTGCAAGTA GAGCTAGCAC AACTTGATTA TTTATTACCT AGATTGCAAG 360  
 GCCATGGTAA AAGCCTTTCT CGTTTAGGTG GCGGTATTGG AACTAGAGGC CCTGGTGAAA 420  
 CGAAGTTAGA GATGGATCGC AGACATATTC GAACTCGTAT GAATGAAATT AAACATCAAT 480  
 15 TGCGGACGGT AGAAGAACAT CGCGAAAGAT ATCGAAATAA AAGAAATCAA AATCAGGTGT 540  
 TTCAAGTAGC TTTAGTTGGT TATACAAATG CTGGTAAATC ATCATGGTTT AATGTTTTAG 600  
 CAAATGAAGA GACGTATGAA AAAGATCAAT TATTTGCAAC GTTAGATCCT AAAACACGAC 660  
 20 AAATTCAAAT AAATGATGGA TTTAATTTAA TTATTTGAGA TACTGTTGGT TTTATACAGA 720  
 AACTACCTAC GACGTTAATT GCAGCTTTTA AATCAACTTT AGAAGAGGCT AAAGGTGCAG 780  
 25 ATTTATTAGT ACATGTCGTA GATAGTAGCC ATCCTGAATA CCGTACGCAG TATGACACAG 840  
 TTAATGATTT AATCAAACAA TTAGATATGA GTCATATTTT TCAAATAGTT ATTTTTAATA 900  
 AAAAGGACTT ATGTGATCAT GCATCAAATC GTCCAGCAAG TGATTGCGCT AATGTTTTTG 960  
 30 TTTCTTCTAA AAATGATGGT GATAAATTAC TTGTTAAGAC GTTATTTATT GATGAAATCA 1020  
 AAAGGCAATT AACTTATTAT GATGAGACAA TTGCGACGAA TAATGCAGAT CGATTATATT 1080  
 TTCTAAACA ACATACATTA GTGACTGAAC TTAAATATGA TGAAATTGAA AATGTTTATC 1140  
 35 GTATAAAGG ATTTAAAAA TAATAAAGG ACGAAATTCA AATGAAAGAT ATAAGTAAGA 1200  
 TAGTAGCTGA CGTCGAATCA ACGTTAGCAC CATATTTTAA AGAAATTGAA GAAACAGCAT 1260  
 40 ATATTAATCA AGAAAAAGTA TTAAATGCAT TTCATCATGT CAAAGCAACC GAAAGTGATC 1320  
 TACAAGGATC AACAGGATAC GGGTATGATG ACTTTGGACG TGATCATTTA G 1371

## (2) INFORMATION FOR SEQ ID NO: 272:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6035 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

55

	CGTAAAcCTA TGCGTTTTAA TATTCTGAAG TTACTTAGTT CATCCTCAGT TTCATCCATT	120
	TGTTTAATAT AAATAATACA TCCAGCTGCT ACTAAAAATG CTAATCCTAA AAATGATGTA	180
5	ACAAATATTA GAATACCGTT AGTAGCATcG ACCTCTTTTT TCATTTTCATC ATACGTGATG	240
	ACTTTGTCTC CAAACTGTTT TGCAATTGCT TGAGCTTTTT CCTTTTGTGA TGTTTGTTTA	300
	ATATCATATC CATAAAAAGT ATGAACGTTA TTTTGTGTTT TCAACTGCTG ATACTTTTCA	360
10	GGACTTACTT CGATGACAGG TGAGTTGAAG CTTAGATTTA AAGGATAAAC CTTACCTTTG	420
	TCTTCTGTG TAcACGGAAA GTTTCATTCT TAGTACCTTT TACTACTAAA TCTTTGTTTA	480
	AATGGATATT AATAATGTTA GGCAGCGATT TTGTATTTGT AATGATGGCA TTGTTGCCTG	540
15	TtAACTTGTT ATTTGCACTT AAAATAGAAT TCGTGCGACC TGAATCACTA CCATTTTCCA	600
	AAGTAATAAC CTGATCATTa ACATTATCTA CAGTAATAGT TTCGTAAGCA TTTTtagAAA	660
20	ATGTAATTTG TTGTGGCTT AGTTTAGTTT CAAATTGTTT AGCATCTTGA GTAGCGACCA	720
	CGTTAAATTC ATTTGGTGCC ATAGATGTAA GGGTTTGATC TGTATTTGAT TTAGATAATG	780
	CCGCAAAACA CAATACAGTT ACTGTAAC TGCAATAAAT TGCAATGATA GTTAAAGACA	840
25	TGGCATTMTT CTTCACTCTG TACATAATAG ACGATGTGAA TACAACATCG GTAATAGATA	900
	CGCGTCCATT TTTTGATTTT TTCAATGTTT TAAAAATAAG TGACACGGAA CTTCTGAAGA	960
	ATAAATAGGC GCCTACAACC GTTAAAAATA AAATGATAAA CGGTGATGTC ATAGCCATAG	1020
30	TTAGTGCTTT GAACGTACCA AACATTCTG TCGCCATATA ATAGCCTAGT GCAATCATAG	1080
	CAATACCTAA TACGCCTGAA ATAACCTCTG CAGTCGTTAC TTTAGCAGTT GTGGCATCAG	1140
	TTTTAATTGA ATCTTTCATC ATTGATAAGA TACTACGTCT TTTTAGAAAT AAAGCACTTT	1200
35	GAAATAAAAT CAGTACATAC GCAATAATTA GCATGAAAAT AGTTAAAACA AGGGCCATAG	1260
	GTTcGAAATG TATCGATAAG TTAATCGATA ACGACATCAA TTTAGATACT ATGGAAAGCA	1320
40	ATAaTTGTGC ACCTGCAATG CCACATAATA CACCGACAAC ACCTGTGATT AAAAATACGA	1380
	TCATTtGTTc AAGTGCTAAC ATTTnCAAAA TGTTTTGTcG TGTTAAACCA ATCaACTGAA	1440
	ATAGCGCAAA TTCACGTGTA CGGCGTTTTA CGrmTAAATG ATTGGCATAC ATTAAAAAGA	1500
45	TGACAATAAT GATAAATAAA AATATTGATC CGACTAAAGC ACCTTTCTTA ATGATGGCCA	1560
	TCGAGTCGTC ATTATTTACA CCTTTAGTAA ACTGTAAGGT TGTA AAACTG AAATATAAGA	1620
	CGATGCTAAA AAATAATGAA AATAAATACA TTGCATAATG TTTTAAGTTT TGTCGTAAGT	1680
50	TTTTGAAAAC GATATGGTTA AATGTCATTT GAGACACCAC CTAATACTGA TTGAAGATGT	1740
	ACAATGTCTT CATAAAAGGC CTGTTTAGAA CGTCCTTCCT GATAAAGTTG TGTATGAATT	1800

55

	ACCATGACAA TAGTTGTATC AAACGATTTA TTCATTTCTT CCAAACGTTG TAATAGGTCA	1920
	TTTGCACTTT TCGAGTCGAG TCGGCCTGTT GGCTCATCTG CAAATATGAT TTGTGGTTTG	1980
5	TGAACAAATG CTCTCGCTGC TGCAGTTCTT TGTGTGTGAC CACCAGATAA TTCGCTAGGG	2040
	TATTTATTTT CTAGGTCATA AATACCTAAT GCTGTCGTGA TCGCTTTATA ATTTTCTTCC	2100
	ATTGTTGCCT TCGACATTTT TTGAACAGAT AAAGGTAACA TAATGTTTTT TTTAACGGTT	2160
10	AATGTCGGCA GAATACTGTA ATCTTGGAAG ATGAAACCTA ATGATTCTTT GCGGAATTTG	2220
	GCAAGTGCTT TTTGATTAAAG TTTATTAAGC TCTGTCCGT TAGCAATCcg cTACCGCTAG	2280
15	AAATTTGGTC AATTGAACTT AGTACATTTA ATAAGGTTGT CTTACCTGAT CCAGAAGGCC	2340
	CCATAATCGC AACGAATTCG CCTTTTTGTA TGTCAAAGTT AATATCTTTA AGTGCTTGAA	2400
	ATGTGTGCTT TTTACCGTAT GTTTTTGAAA CATGTgCACT GATAATATCG TCATAGTCTC	2460
20	ACTCCTTtTG TATTTAATTT CATTTTAAAT AATGTTTGGG GTAGTAGCCT TTATCTAAAC	2520
	TTACAATTCA ATGAATGAAC CTTACAGAGT TGAARcTAT CGCTACTTAG TAGATTTTTG	2580
	AGTGAGGATA CAGATTCATC GTACATATTA GACAAAAGCA ATGGTGCTTT CTAAGTGATG	2640
25	ATGTTTGTGT AAATTGAGAA AAGGGAATTT AATTATTGTA TAATAAATTT TTTGTAAAAA	2700
	TTAAAAGAGG GTTTTATTTG AAAGGAATTG ATTGTTATGG AAAAAGGAAA TCAAGGTATT	2760
	AAATGGTCTA GTTTAATAAT GGGTGATTA TTATTAATGT TGGCAGTCGT TATTTTTACA	2820
30	TTTCCAATTG AAAATTTTTA TGCTATTACC TGGTTGATTG GACTGTTTGT ATTAATTAAC	2880
	GGTGTGATTG AAATCGTTTA CCGTAGAAAA GCAAAAGCTT TAGTAGGTGG TAACCAAAAT	2940
	TGGATCTGT TTATGGGGAT TGTAATATT CTATTTGGTC TATTAGTTAT TTTAATGTT	3000
35	GGCGCAAGTT CAGCATTCTT TATTTATATG TTTGCTTTTT GGTTTATTTT TAGTTCTATC	3060
	TCTGGATTAT TTACGTTTTT GGGTAGTGGT AGCTTAAAC TAATTCAGT GATTTTTAAT	3120
40	TTATTAGGTA TTGTTTTCGG TGTCATTTTA TTATTTAATC CATTAATGGG TATCGTCTTT	3180
	ATTCGACGA TGATTGCTAT TGCATTTGTA TTCGTAGGTG TCATTTATGT TGTAGATGCA	3240
	CTTGCTTAAG TAAATGAAG CGTTCAAAA GAAGGTGTG ACATGAAGTT TGTGTCATAT	3300
45	CCTTTTGTG GTGTTTATGA AGCATAAAAA AGGGGCGCTA CCTACAATAA GTAAGATACG	3360
	CCCATATTTT TATATTTTAC TATTATTGTT TTTCAATACG ATTAATAGTT ACATTTAGTC	3420
	CAAAATATTT TTCTAAAAAA TGTTTATAGT TATCTTTAGT GACATCAAAT TTTTCTGAGC	3480
50	TACCATTCTT TGTTAAAGTT AAATGATTTT CAGACATTGT AGCACGGCCA AATGATTGTG	3540
	GCATTGTAAT TAATAAATGC TGTACAAATA TTGAATCTGG ATGCGTTTGA TTATATTGCA	3600

55

EP 0 786 519 A2

	AATGATCATT TTCGAATTTT TGAACATAGA AAATATCCTT GTCTTCGTTG TTAAAAATAG	3720
	CGCGGAATGT ACCACTGATA TCAGTAATTG GTTGTGTATG CTCAGATGAA GTAATAGGAA	3780
5	TGGCATGTAG AGGTAAGTCT CCAAAGCCAA CATCAGTTAC ATAGAATACA TCATTTATAG	3840
	AAACAACAAG TGAAGCATGT GAACCGTTCA GACTACGACC GCCACCGGGW GTGTGAATAG	3900
	TAGCTGACAT TAATTCAGGA TTAAATCCTT TTTGTTGTAA ATAGGCTTTG AAAAAATGTAT	3960
10	TTAATTCATA ACAAAAACCA CCACGTTTAT CATGAACAAT TTTATTAAAA AGTGCATCGA	4020
	TATTTATAGA TATCGGCTTA CTATTTTGAA CATCAATATT TTCAAAAGGT ACAGTTAACA	4080
15	TAAAACGTGT TGCATAATAA TTTAATGCTT CAATACTCGG TCGATTATAA CGAGATGAAT	4140
	CAATTTGTAA ATAATTCTCT AACTTCGCAA TATTCATAAG CATAGCGCCT CCTGTATTAA	4200
	AGATTATAAT TAAATTTTAA ACAGAAATAC TGAAATTTTA AATTCGAAAG CATTGAATTT	4260
20	TGGATAAATA CATTTTAAAT AGAAAAATAC GCTCTCAAAA TGAAGTCATC TCTAAAAGAA	4320
	ACGATTTAAA GATGACTACT GAGAGCGTAg CATAATGGAA GAAGTGTGCA GGGTGTCTAA	4380
	AAATGCAACA ATACAAAGGT AGTTGCAAGA CAAGTTGCCT TATCTAGACC ATTGTGTTC	4440
25	TATGCGACCA AACTTCCAAA TTAACTTGA AATAAGCCAA GTAATTAAAA ATAATGCAAC	4500
	TAAATATAG CCTAAATAAT CAAATTCGAT CGAACCAATG AATGCCCAAA ACGCACCATG	4560
	TAAATCTAAC TTATCAGCAA GAATTTGTAG CAATTCAATC ATCCCAATCA CTAATGCTGC	4620
30	CATGACTGAT ATCGCAGTAA TCGTTATATT GTAATAGATT TTGCGAATAG GATTGAAGAA	4680
	TGCCCAATTA TAGGCATACT TCATTACAAC ACCATCTAAT GTATCCAATA AACTCATACC	4740
	TGATGCGAAT AAAATTGGTA AAGATAAGAT TCCGATAAAT GAAATGGCTT GTTGTGATGC	4800
35	GCCTGAAGAA AGAGCGAGTA ACGCAATTTT ACTAGCTGTA TCAAAACCAA GTCCAAATAA	4860
	AAAGCCAAGT GGCAATACGT GCCAACTACG CGTGATTAAAT TTGAAATAAG GTCCTACAAA	4920
40	TCGAGAAACC AATCCTCTAG ATTCAAGTAA TGCATCGACT TCAGCTTCTT CAATGTGTTC	4980
	ACGACGTAAT TTAGCGAACA AGTTAATTAA AGAGATTAAA ATAATTAGAT TCAACACACC	5040
	GATAAGCACT AAAAAGAAAC CTGAACTAG TGTACCAATC GTTCCACCAA TATCTTGAA	5100
45	ATGCGGTAAT TCATCTTTAG CCCATTTTAC AGATACCCCT AAAAAACAG CCATTAAAAA	5160
	TACGACAGAT GAATGTCCAA TTGAAAAATA GAAACCCACA CCAGATGGAT CTTTGCCTTG	5220
	CTGTAATAAT TTGCGaACCG TATTATCTAT TGCAGCAATG TGATCTGCAT CAAATGCATG	5280
50	ACGCAAACCT AATGTATATG CAAGAATCCC CaTACCAAAT AAGATATGAT GGTCTTTTCC	5340
	AGCAATCCAT AAAAACTAA ACCCAATAAC GTGTAACAAA ATGACAATAG CTATGTATGG	5400

55

ATATTTAATC ATACTGTATG TTCAATGGGC ACTCTAGTAA TAAGTGTTCA TATAACAAAA 5520  
 ATGTTATGCC AAATTATTTG TTATATAAAA ATATACATGT AACCACAAAA GATTTTTTGC 5580  
 5 GATATATATA ATTTGATAAA TTAACCAACA ACAATGTAAG ATGTCACCTT GCTTAACTTG 5640  
 GCATCCTTTT TATGATTTTC AAATTCAAAA AAATGAGCAA AATGAATCTC TTTACcAGTT 5700  
 TTTAATATTT CaATACCATG CATGGAACCT AAGCACCCAT GTGTGATGCT GGAATGGATA 5760  
 10 TTGAGACTAG CAACCTGATT GTAATGATTA GATAGTTCTT GAATTAATAT TTGAGGTCCG 5820  
 TATATGTCAA AGCGGCCAGG GACAGACCAA ATAAATTCTG TTGTAACCAG TGAACGTAAT 5880  
 AATTCAATAT CTAATGCTGC TGTAACAACT ATAAATCTA TCATTTGTTG ACGTTTAGGC 5940  
 15 GCATGATTGC ATGACACATC TCCTGTTAAC TTAAGGTA ATGATGACTG AACTTCCGTT 6000  
 TTAAGTGTGTA GTTGGTGCTG AAATAAGCT TGTTC 6035

20 (2) INFORMATION FOR SEQ ID NO: 273:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1039 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:  
 30 TTTTGAACAG CCATATTTAT TCACCCTCAA CATCATTATA ATGGTATTAG TCGCATTACC 60  
 TTCCTTGTT TTAGCTATAT ATGATTATAT GAGTTTAGA ATTTCTTCTG CTATTTTACA 120  
 ATTTCTAGGG GCTATCTCTT GGTTCCTTTT ATCATTGATA TTATCGCTCA CACAATTTAC 180  
 35 ACCTTTTACA TTAGCGTCAT TTATAACTTC AATTATTTTG TTCACAAGCA CAATTATCAC 240  
 ATTAGCCATT GGTGGTAAGT CTGTTGAAAA GAATGATTCC CCTTAAATTC CAAATGAAAA 300  
 AAAGGTTCTG AAGGCCGCTA TAAACACAG TTTTTCAGAA CCTCTATACT TCTATTCAAT 360  
 40 GATATATGGT TTGCAATTTT CTACCTTTAA ATCCACAGCT TCTGCCCTTG AAACCTTGTT 420  
 AAAATAAACC ATCAAACAAC GAATGACAAC TTGATGTGCA ACAATGACAA TATCATCTTT 480  
 45 TTGTGTATCT TCATTGACAA CATGATTCAT AAAATGTTCT ACGCGTTGAT ATACATCTTC 540  
 ATAACTTTCT CCTTCAGGCG CTTTTGTGA AAAATATGA CGAAAGTCTT TAAAGTTTGG 600  
 ATCATTGAAA TATTTTTCAT ATTTCCGATT CGCACTGATT TCATCTTTAT ATTCACCCTC 660  
 50 AAATACGCCA AGTGAACGTT CTCTAATAG AGGGGTAGTC GTTGATGCAA TGTCATATGG 720  
 AAAAATATGT TCAAACGTTT GCTGTGTTCT TAATAAGTCT GAAACATATA CATGTTTAAT 780

55

CGGCACATCT AATTGTCCAC AAAAATAAGA TCGAAAATGT TTATTATCAT AATTCGATTT 900  
 TGATTCGCCA TGTCTAACTA AATAAATCGT CATAATATTA CTCCTTACCT TATGTATTTC 960  
 5 ATATCTACCA TAACACTTTG ACTACTAATT CGATATCAAT CTTAATATTC TATTCTAAAA 1020  
 AAAGAATTAA TTCATATnT 1039

## (2) INFORMATION FOR SEQ ID NO: 274:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1496 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

GAGAGAATnT GCAATTAGTT ATTCAATTAG TTGATTTAAG ACATGATCCA ACACAAGATG 60  
 20 ATATCTTAAT GTACAATTAT TTGAAACATT TTGATATTCC TACTTTAGTT ATATGCACTA 120  
 ArGAaGACAA AATTCCaAAA GGTAAGGtYc AAAAGCATAT TAAAAATATT AAGACACAAT 180  
 25 TAGATATGGA CCCAGACGAT ACAATTGTAA GTTATTCATC AATTCAAAAT AATAmCAAC 240  
 AACAAATATG GAATTTAATT GAACCGTATA TTTCATAGTT TTTGTACGTC AAAACTTATA 300  
 CAAAAATTTT AAAAATAATG TAAGCACGAA ACTTTTAATT AGTACACAAT TGATAACATT 360  
 30 TTTCAACGTT CATCATTTTG TCAAAAACtC AAAAGTAAAT TAGAAAGATT ATAATTTATT 420  
 TAAGCATCGT ACTTAATTGG ATTTTAAATT ATGTTATAAT ATTTGTATTG TTAGTATATA 480  
 TGGGGGCTTT TCAAATGCAT TTTATTGCAA TTAGTATAAA TCATCGCACA GCTGATGTgC 540  
 35 ACTAAGAGAG CAAGTTACTT TTAGAGATGA TGCCTTACGA ATTGCCCATG AAGATTTATA 600  
 TGAAACTAAA TCTAyTTTAG AAAATGgTCA TATTaTCAAC ATGTAATCGA ACTGAAGTAT 660  
 ATGCTGTTGT TGATCAAATT CACACAGGTC GTTACTATAT TCAACGATTT CTAGCTCGTG 720  
 40 CATTTGGATT TGAAGTAGAT GATATTAAAG CAATGTCAGA AGTAAAAGTG GGGGACGAAG 780  
 CaGTAGAACA TTTATTGCGT GTCACtTCTG GTTTAGATTc AATCGTACTT GGAGAAACTC 840  
 45 AAATTTTAGG TCAAATAAGA GATGCATTTT TCTTAGCGCA AAGCACAGGT ACGACAGGrA 900  
 CAATTTTAA TCATCTATTT AAACAGGCAA TTACTTTTGC AAAAAGAGCA CATAATGAAA 960  
 CAGATATAGC TGATAATGCT GTAAGTGTGT CTTATGCTGC GGTcGAGTTG GCGAAAAAG 1020  
 50 TATTTGGCAA ATTGAAAAGT AAGCAAGCTA TCATTATTGG TGCAGGGGAA ATGAGTGAAT 1080  
 TATCACTATT AAATCTTCTT GGTTCTGGAA TTACTGaTAT TACAGTAGTA AATAGAACAA 1140

TACCAAATTT ACTTGAAAGT GCAGATATTG TGATTAGTTC AACGAGTGCA CAATCTTATA 1260  
 TCATTACAAA TGAATGATA GAAAGAATTG CAGAAAATAG AAAGCAAGAT TCACTAGTAT 1320  
 5 TGATTGATAT TGCAGTTCCT CGAGATATTG AACCTGGTAT TAGTGCCATC ACAAACATCT 1380  
 TTAATTATGA TGTGATGAC TTAAGGTT TAGTTGATGC AAACCTACGT GAGCGACAAT 1440  
 TAGCGGCTGC AACAAATTCG GAACAAATTC CTACAGAAAT ACATGCACAC AATGAG 1496

10 (2) INFORMATION FOR SEQ ID NO: 275:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4826 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

CTTGATTTT TCCCTTAGT ATTTCCAtT TGanTGTCGC AGCTTCTAAA TCCTGCTTTG 60  
 GTTCTCTAGT GAACTTCATA ATTAAAGCAG CTACAACGAA TGATACAAGT GCAGCAAGGA 120  
 25 AGACACCGAG TAACATGTGC AAGAATTCAC CTCTAGGTGC ATTTAAACAG TAAACTATAA 180  
 ATGAACCTGG TGACGCGGGA CTTTAAATC CAAATCCTGT TGCTTGATAA GTTGCAACAC 240  
 CAGTCATTCC ACCTAAAATA ACAGCGATAA ATAATAAAGG ACGCATTAAAT ACATATGGGA 300  
 30 AATAAATTC ATGAATACCA CCTAAGAAGT GGATAATTCC AGCACCATAT GACGTTGCTT 360  
 TTGCAGTGCC TTTTCCAAAA ATCATATAAG CAAGTAAGAT ACCTAACCT GGTCCAGGGT 420  
 TAGATTCAAT TGTGTATAAA ATTGATTGAC CAGCTTTTGC AGCTTGATCT GCACCAAGCG 480  
 35 GTGTGAATAC ACCATGGTTA ATCGCATTGT TAAAAATAC AATTTTGTCA GGCTCTACTA 540  
 AAATACTTAC AAGTGAAGT AGGTGTGCAT GTACTAATGC TTCAACTGCC ACTGATAAAA 600  
 TATGCATAAT AAATTCATA AGTGGTGCTA AAATTTTAAA TCCTGCAATC GTCATGATAA 660  
 40 ATCCTAAAAT ACCAGCAGAA AAGTTATTAA ATAACATTC AAAACCTTGC GCGTTCTAG 720  
 GTTGAATCAA TTGGTCGGTC TTCTTCATTA ACCAACCAAC AAGTGGACCC ATAATCATTG 780  
 45 CACCAAGTAA CATTGGTGTA TCAGGTAATG CAACGATGAC CCCCATAGTT GCTGTTGCTG 840  
 CGATGATACC ACCACGTAAA TCATAAATTA AACGACCACC ACTAAATGCG ATCAATAATG 900  
 GGATTAAATA AGTAATCATT GGTCTGCTA AAGTAGCTAA ATCTTTGTTA GGTAACCATC 960  
 50 CATTATCTAT AAAAATGGCC GCGATAAAAC CCCAAGCGAT GAAAGCGCCA ATGTTTGGCA 1020  
 TGATCATACT ACTTAAGAAT GATCCAAATG CTTGAACACG ACGACCAATT CCTTTTTTCT 1080

55



## EP 0 786 519 A2

	GAGAGGTTAC	TTGTTACTCA	ATATAACAA	AAATCAACTT	TGTCAAAATA	AATGTGACAA	1200
	AATTAAATAA	AGTGTCATCA	ATGTGACAGT	ATAGATATTT	TGAAAAAGTA	AAACAAAAAA	1260
5	ATTGTTTTAG	GATTTTTAAA	ATTTTATTGT	GAAAAATTTT	GCAAAACAAA	ACAACACCGT	1320
	GTACAATAAT	GATTAATGGA	AAGGGGGAAA	GTTCCGGCAGT	ACAGTTAAAG	CGCCTGTGCA	1380
	AATAAATATT	TGTATTTGAA	GATTAAAGGT	TAATATATGA	GTGGCCTTTA	TAGAGTGCAA	1440
10	TATATGTATT	TGTAGACGAG	GAGGATAGTG	ATCGAATAGA	TCGGCGGATG	CTATCCCGGA	1500
	TGTGGCTCAT	TCGTTAGCTT	ATTAAGTAAA	ACATTAGGGT	GACTTAATGG	ACAAAGTTAA	1560
	TAAGATCGCC	AGAAATTGAA	TATAAAAAAT	ATTAATATGG	AAAGTACAGT	GTGAGCAATT	1620
15	TGTATAGTTG	TAAAAATAAC	TATGCTTAAT	TTGTTATGGA	TGAATGCGAT	GATAGCATGT	1680
	TCCTATTTAT	ATTATGAAAG	CAGATTGTCA	ATCTAAATTA	TCGGCAATAA	ATCATAATTT	1740
20	ACGCGTACTA	TTCCAATATG	GAGGAAAATG	TCGTTATGTG	TGGAATTGTT	GGTTATATTG	1800
	GCTATGaTAA	TGCCAAAGAA	TTATTATTAA	AAGGTTTAGA	AAAATTAGAA	TACAGAGGTT	1860
	ATGACTCTGC	AGGTATCGCA	GTAGTAAATG	ATGATAATAC	AACTGTATTT	AAAGAAAAAG	1920
25	GTCGTATTGC	AGAATTACGT	AAAGTTGCTG	ATAGTAGCGA	TTTTGATGGA	CCTGTTGGAA	1980
	TCGGTCACAC	ACGTTGGGCA	ACACACGGTG	TACCGAATCA	TGAAAACCTCT	CATCCACATC	2040
	AATCATCAAA	TGGCCGTTTT	ACTCTAGTTC	ATAACGGTGT	TATTGAAAAC	TATGAAGAGT	2100
30	TAAAAGGTGA	ATACTTACAA	GGTGTATCAT	TCATTTTACA	AACAGATACA	GAAGTTATCG	2160
	TTCAATTAGT	TGAATACTTT	TCAAATCAAG	GACTTTCAAC	TGAAGAAGCA	TTTACAAAAG	2220
	TTGTGTCATT	ATTACATGGT	TCATATGCAT	TAGGTTTATT	AGATGCTGAA	GACAAAGACA	2280
35	CAATCTATGT	TGCTAAAAAT	AAATCACCAT	TATTATTAGG	TGTTGGTGAA	GGTTTCAATG	2340
	TTATGCGCATC	AGACGCACTT	GCAATGTTAC	AAGTGACAAG	CGAATATAAA	GAAATCCATG	2400
40	ACCATGAAAT	CGTTATTGTT	AAAAAAGATG	AAGTTATTAT	TAAAGATGCA	GATGGAAACG	2460
	TTGTAGAACG	TGATTCATAT	ATTGCTGAAA	TTGATGCATC	AGATGCTGAA	AAAGGTGTTT	2520
	ATGCACACTA	CATGTTAAAA	GAAATTCATG	AACAACCAGC	AGTAATGCGT	CGTATTATTC	2580
45	AAGAATATCA	AGATGCAGAA	GGTAACTTGA	AAATTGATCA	AGACATCATC	AATGATGTTA	2640
	AAGAAGCAGA	CCGCATTTAC	GTTATTGCAG	CAGGTACAAG	CTACCATGCA	GGTTTAGTAG	2700
	GTAAAGAATT	TTTAGAAAAA	TGGGCTGGCG	TACCAACTGA	AGTACACGTT	GCATCAGAGT	2760
50	TTGTCTACAA	CATGCCATTA	TTATCTGAAA	AACCATTGTT	CGTTTATATT	TCTCAATCAG	2820
	GTGAAACTGC	AGATAGCCGC	GCCGTATTAG	TTGAAACTAA	TAAATTAGGT	CATAAATCAT	2880

55

	TACACGCGGG TCCTGAAATC GCAGTTGCAT CTACAAAAGC ATATACTGCA CAAATTGCAG	3000
	TATTATCAAT CTTGTCTCAA ATCGTTGCAA AAGAGCaTGG TCGTGAAGCA GATATTGATT	3060
5	TATTGAGAGA ATTAGCAAAA GTAACAACAG CAATAGAAGC AATTGTTGAC GATGCACCAA	3120
	TTATGGAACA AATTGCTACA GATTTCCTTAG AAACAACACG CAATGCATTC TTTATCGGAC	3180
	GTACTATTGA CTATAACGTA AGTTTAGAAG GTGCGTTAAA ACTTAAAGAA ATTTCTTACA	3240
10	TTCaAGCAGA AGGTTTTGCT GGTGGAGAAC TTAACATGG TACAATTGCC TTAATCGAAG	3300
	AAGGTACACC AGTTGTAGGT TTAGCAACAC AAGAGAAAGT TAATTATCA ATTCGTGGTA	3360
	ACGTTAAAGA GGTAGTAGCA CGTGGTGAC ATCCATGTAT TATTTCTATG GAGGGTCTTG	3420
15	AAAAAGAAGG CGACACTTAT GTCATTCCTC ATGTACATGA ATTGTTAACG CCATTAGTAT	3480
	CAGTGGTTGC ATTACAATTA ATTCATACT ATGCAGCATT ACACAGAGAT TTAGATGTTG	3540
20	ATAAACCACG TAACCTTGCT AAATCAGTTA CTGTGGAATA ATTCACTTTT TTAGAATCAA	3600
	TCATGTATTA AAATTAAAGT ATATGGCACC CTTTTAGATT AATCGACTAG AAGGGTGCTT	3660
	TTTTAGGTCG ACTTaGCITT TACTTCATCT TAATTTGGCA GAAATGCGTa AAAATGAAGT	3720
25	GTTTTATTTA TTTAAATAGT CTGACAATTA AGGGTGTTAT GTTAATATGA TTTTATGAGA	3780
	AGTATGGAGT AGCAATAAAG GGGTGACCTC GCATGTTAAT TCAATTAGAT CAAATTGGGC	3840
	GAATGAAGCA AGGAAAAACA ATTTTAAAAA AGATTTCTTG GCAAATGCT AAAGGTGATA	3900
30	AATGGATATT ATATGGGTTG AATGGTGCTG GCAAGACAAC ACTTCTAAAT ATTTTAAATG	3960
	CGTATGAGCC TGCAACATCT GGAACGTGA ACCTTTTCGG TAAATGCCA GGCAAGGTAG	4020
	GGTATTCTGC AGAGACTGTA CGACAACATA TAGGTTTTGT ATCTCATAGT TTA CTGGA	4080
35	AGTTTCAAGA GGGTGAAAGA GTAATCGATG TGGTGATAAG CGGTGCCTTT AAATCAATTG	4140
	GTGTTTATCA AGATATTGAT GATGAGATAC GTAATGAAGC ACATCAATTA CTTAAATTAG	4200
40	TTGGAATGTC TGCTAAAGCG CAACAATATA TTGGTTATTT ATCTACCGGT GAAAAACAAC	4260
	GAGTGATGAT TGCACGAGCT TTAATGGGGC AACCCAGGT TTTAATTTTA GATGAGCCAG	4320
	CAGCTGGTTT AGACTTTATT GCACGAGAAT CGTTGTTAAG TATACTTGAC TCATTGTCAG	4380
45	ATTCATATCC AACGCTTGCG ATGATTTATG TGACGCACTT TATTGAAGAA ATA ACTGCTA	4440
	ACTTTTCCAA AATTTTACTG CTA AAAGATG GCCAAAGTAT TCAACAAGGC GCTGTAGAAG	4500
	ACATATTAAC TTCTGAAAAC ATGTCACGAT TTTTCCAGAA AAATGTAGCA GTTCAAAGAT	4560
50	GGAATAATCG ATTTTCTATG GCAATGTTAG AGTAAATATT TTGCAAATAA TAAGTAATAA	4620
	TGACAAAATT TAATTAAGAT AAAATGGACA GTGGAGGGCA ATATGGATAA CGTAAAAGCA	4680

55

AAAGATGTCA TTAATCAATT GAGAGAGAAA GGATATAAAG TATTTTGGC AACAGGACGT 4800  
TCGCATTCTG AAAATACATC AACTTG 4826

5

(2) INFORMATION FOR SEQ ID NO: 276:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4846 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

15

GAATAAAAAG TAAATTTACT CGCCTTTGTT ACCTTTTACT TTATCAATaA AATCAGTTGC 60  
TTTTTCTTTT GCATTTTCAA CGAATTCTTT CGCTTTACCA GAAGCTTTAT CTTCTTTACC 120  
TTCGTTTTCT AAATTTTAT TATCAGTAAC ATTACCTACT GTTTCTTTAA CATTACCTTT 180  
TGCTTGTTCA AATTTACTTT CGTCTGCCAT AATAAATGCC TCCTCGGAAT AATTAAATGT 240  
TATATATAAT ACTTACCCAC TGAAAAATTA TCTAAACATT TTAATTAAAT AATTTTGTAT 300  
ATTGATTTGA CGTCATTTTA TAACTAGCGA AATAGATTCA TCATTAACCT GAGGGAGTGG 360  
GACTGAAATA ATAAAGAATC ACTAATGATT TATGATGTAT TAGTCACTAG CCATGTGAAA 420  
TTAAAAATAA GAATAAATGA GTAGCACGCA TGCATATAGG ATTTTACTTT ATCCGTAATA 480  
GCATCTCATT CCTAAATATC ATATAAATAC CTGTTTAAAT TAAAAAGCCC AGCAACATCA 540  
CGTTACTGAG CCATTAATAT GATTTATTTA GCAGGAATAA CTGCACCATT GTATTTTCA 600  
TTAATGAAGT CTTGAATATC TTTAGATTGT AATACTTCAA TTAATGCTTT GATTTTCTTA 660  
TCATCTTGAT GTCCTTCTTT AACAGCAATT AAGTTTGCAT AAGGATTATC TTTCGCACTT 720  
TCTACAGCAA TAGAATCTTT TTTAGGATTT AGTTTTTGTG CGATTGCAAA GTTCGAATTA 780  
ATGATAACAG CATCAGCGTC TTCATTTTGA TAAATTTTAG GTAAGAATTC TGCTGATTGT 840  
TTATTATTAA ACTTAATATC TTTTTTATTC TCTGTAATAT CACTAACTT AGCATCTTCA 900  
ATTTTACGCG CTTTTTTGAT TTTAATTAAA CCTGCATCAA CGAAGAATTT TAAGAAACGT 960  
CCTTGTTTCTAG CTGGATTATT AGACACATAG ACTGTTGCAC CTTTGGTAA TTCTTTTAAA 1020  
CTTTTATACT TTTTAGAGTA TACAGCCATA GGTCTAAGT GAACATCACC GGCCTTACG 1080  
ATTTTGTAAC CTTTATCCTT TTTCTCTGTG TTTAAATATG GTGTATGTTG GAAATAGTTT 1140  
GCGTCAATTT CACCTTTGTC TAGTAATTTA TTAGGTGTAG TGTAATCGTT AATTGTTTTA 1200  
ATATCTAGTT CATAACCTTT TTTCTCTAAT AATGGTTTTG CTTTTTCTAA AATTCAGCA 1260

55

	TTACCGCAAG CTGCTAATAC AACTGCAAAT GTTAATACTA AAATAAGACC AAATAATTTT	1380
	TTCATAAAAT GAAACCCCCA ATTTATCGTT TATCAAGTTT ATTTGTAAGC CAATCCCCAA	1440
5	TGAATTGGAT TATAAATACA ATAATTAAAA TAAAACTGT TGATACTAAA ATGACATCAT	1500
	TTTGATTTCG AGTGAAACCT GTTAAGTATG CTAAATTTCC TAAACCACCG GCACCAATTA	1560
	CACCTGCAAC TGCTGTTGAA CCAACTAAAG CGATTGCTGT AACTGTAATG CCAGACACTA	1620
10	GCGCTGGCAT AGCTTCAGGT AAAAGGACTT TACGAATTAC TGTCCAAGTA TTAGCGCCCA	1680
	TTGACCAAGC CGCTTCGATG ACACCTTTAT CAATTTCTTT AAAAGCAATT TCTACGAGCC	1740
	TTGCATAAAA CGGTGctGCG CCAATGATCA AGGCTGGTAA CGCACCTGTC GGACCACTTA	1800
15	TCGTTCCAAG TATCAAACCT GTAAATGGAA TTAATAATAA AATTAAAATA ATAAATGGTA	1860
	TCGCTCTAAA TAAGTTAACA ATGAAAGAAA CGATAGAATA AAATAACCTT GCACCGATAG	1920
20	ACTTACCTTT AGCAGACAAG AATAATAACA CACCTAAAAT AAGACCAAGT ATAAATGCAA	1980
	ATATAGTTGA GACGACTGTC ATGTATAGTG TTTCGACTAT TGCAGTCCAA ACTTCTGGCC	2040
	ACTGAATATT AGGCATTGTA ATCATTTCAT TTATAATTC ACTAAATGAT TTACCCATGT	2100
25	CTTAACACCT CCATTTTAAC TTGTCGCTCA ATTAACCTT TTTCGAATTT TCCGAAATCT	2160
	ACACTTGAAA TATATGGAAT ATGCAGAACT AAAAAGCCGA CTGTTCCATT TTTTGTATTT	2220
	TTAATATTTG CTCTCTAAAAT ATTAATTTTA ATATCATAGG CAGTTGATAG ACTCGATACA	2280
30	ATAGGCTCGG TTGTTGTTGA ACCAGCGAAA ACTAATCTAA CGATATATGC ATCTTTTTCT	2340
	AATGGCTCTA ATTCTGTAA AGATGTTTCG AAATCATCAT TTAAATCGTC TTTCACAAAT	2400
	CGTTTTGTCA CAGTGTGTTG CGGATTTTCA AAAACCTGTG TCACCGGTCC TTGTTCTATC	2460
35	ACTTTACCAC TTTCCATAAC TGCAACTTCA TCACAAATAC GACGAATGAC ATGCATTTCA	2520
	TGCGTAATTA GTACAATTGT TAAATTTTGT TGTCTCTAA TTTTAGTAG TAGATCTAAA	2580
	ATTTTATCTG TTGTTTTCGG ATCAAGTGCA CTTGTTGCCT CATCACAAAG CAAGACCGTT	2640
40	GGATCATTaG TAACGCTCGT GCAATCCCAA CACGTTGCTT TTGTCCACCT GATAACTCTG	2700
	ATGGATAAGC CTTTTCTCTA CCTTTTAAAC CGACGAGTTC GACAAGTTCT AATGCTTTTT	2760
45	GCTTAGCTCT CTTCTAGGG ACACCTGCAA TTTCAAGCGG AACATAATA TTTTTTAAAC	2820
	CAGTCCTTGA CCATAACAAA TTAAATGTT GGAAGATCAT ACTTACTTTT TGTCTTTTTG	2880
	CTCTTAATCC ATTTTGGAC AATTGACCTA TATGGTCTCC ATCTATAATA ACTTCACCTG	2940
50	ATGTAGGCGC TTCTAAATGA TTAAACATTC GAATCAAAGT ACTTTTTCCT GCTCCAGAAA	3000
	AACCAATGAC GCCATAAATC GATCCTGCTC GAATCGATAA ATTAACGTGA TCTACAGCAA	3060

55

	TTCCTCCCTG TGTGCTTAA TAAAATAAAA AATGCTTTCT CAATATCGAT AGAAAAATTG	3180
	AGAAAGCAAT AGTAGTATTG TTTCTCTCAT CTTCAAAAGT TAAACTTTA TGTGAATTGG	3240
5	CACCATTCT ATATAAGACG GTTGCCGGGC TTCGTAGGGC ACATCCCTCC ACCACTCTCG	3300
	ATAAGAGTTT ACGCATCATT TAATTGTAT TAATCCTAAC ACCTTAGTAA AATTCGTCA	3360
	ATAACTATTT TAAATTTTCT AACAAATCAG TCACCGATT AAATGCATAA ATTCGTTTTA	3420
10	CTTCTTTATC TTTATTCATC AACAATAAAA TCGGCGTAGA CATGATTGTC ATATCTTTAC	3480
	AAACTGAGG ATAAAAGTTT AAATCTATTT TCAATAATGG TAACTGCAAT ATTTCAATTAG	3540
15	CAATGTCTAA CATTCTTTCT GAAACCTTAC AAGTACCACA CGTTGGTGTA TAACCAAAGA	3600
	TTAAATGTTT GTCTTCCTCA TAAATGTAG TTACATCTTT GATGTCTAAT GAATTATTCA	3660
	TTTACTAAAA CTAACCTTTC ATTATTTATA TTCGGTAAAA GAGGTGTTTC TTTCTTACAA	3720
20	GTAAAGCCAT GTTTTGAAAG TACATGCGCC AAATATTGTT TGGGGCAATT CGCAACTTGA	3780
	CAGTAAGTTT TATCAATAAA TATATGTTCA CTTTCACTCA AATAACGTTT AAACCAATTT	3840
	CTAATTCGAT CTCCTTCGTC ATCAGAATCG GCTAATACAA AAACCTGTTT ATCATAAGT	3900
25	GATTCTATCA TATCATCAAG CTTATCTATA CTCATTGTTT CATGAGTACA AATAATATTG	3960
	ACTGGTTCTG CAATAACCTG TTGCACCCTT TTTTATCAG ATTTTCCTTC AACAAATTATC	4020
	ACTTTATTTA CAATAGCCAT CATCATCACC CTTTAAAATC AATAAACATC TGTCACGTGA	4080
30	TCATTTTACA AAATTGGTAT GAATAAAACA TAAATCACAA AAAATTTAAA CTAGCTTAAT	4140
	ATAATAATTA CAAACTCAAT GTTTGACTAG CTGGAACATT TAACATAAGC AGACAAAGGC	4200
	TAAGTCAAAA ATCAACATCC TAAAATCTAC AATGTATAT TAACAATAGT TAACCAAAG	4260
35	AAAATACACC TATAACAAAC TTTTCAATTA TAGCGGGGCC CCAACACAGA AGCTGATGGT	4320
	AAGT <sup>1</sup> AGCTT ACAATAATGT GCAAGTTGGC GGGGCCCCAA CATAAAGAAA TACTTTTCT	4380
40	TTAGAAATTA GTATTTCTTA TGCATGAGTT TTA <sup>2</sup> CTATGT ATTCCTATTT TTAAATACAC	4440
	ATTAGCTGTG GCTTATGAAA ACAGGCTGGG ACATAAATCA ATGTTCTATG CTCTACGAAG	4500
	TTATATTGGC AGTAGTTGAC TGAACGAAAA TGC <sup>3</sup> GTTGTA ACAAGCTTTT TTCAATTCTA	4560
45	GTCAGGGGCC CCAACACAGA GAATTTTCGAA AAGAAATTC <sup>4</sup> ACAGGCAATG CAAGTTGGGG	4620
	ATGGGCCCCA ACAAGAGAA ATTGGATTCC CAATTCTAC AGACAATGCA AGTTGGGGTG	4680
	GGACGACGAA ATAAATTTTG CGAAAATATC ATTTATGTCC CACTCCCTAG ATTGATCTAT	4740
50	AGATACTACA CTTATTAAAG TAATATATTT TTATGATTCT CTTAGCTGCA ATCCCATGAA	4800
	TACATGTAAT CATCAAACTT CATAGCCTCA AGGTCAGTAG ATTTCA	4846

55

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1843 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

AACAAAGACA CAATCGAACA TGAACCATCA GTAAAAGCTG AAGATATATC AAAAAAGGAG 60  
 GATACACCAA AAGAAGTAGC TGATGTTGCT GAAGTTCAGC CGAAATCGTC AGTCACTCAT 120  
 AACGCAGAGA CACCTAAGGT TAGAAAAGCT CGTCTGTTG ATGAAGGCTC TTTTGATATT 180  
 ACAAGAGATT CTAAAAATGT AGTTGAATCT ACCCCAATTA CAATTCAAGG TAAAGAACAT 240  
 TTTGAAGGTT ACGGAAGTGT TGATATACAA AAAAAACCAA CAGATTTAGG GGTATCAGAG 300  
 GTAACCAGGT TTAATGTTGG TAATGAAAGT AATGGTTTGA TAGGAGCTTT ACAATTAAAA 360  
 AATAAAATAG ATTTTAGTAA GGATTTCAAT TTTAAAGTTA GAGTGGCAAA TAACCATCAA 420  
 TCAAATACCA CAGGTGCTGA TGGTTGGGGG TTCTTATTTA GTAAAGGAAA TGCAGAAGAA 480  
 TATTAACTA ATGGTGAAT CCTTGGGGAT AAAGGTCTGG TAAATTCAGG CGGATTTAAA 540  
 ATTGATACTG GATACATTTA TACAAGTTCC ATGGACAAAA CTGAAAAGCA AGCTGGACAA 600  
 GGTATAGAG GATACGGAGC TTTTGTGAAA AATGACAGTT CTGGTAATTC ACAAATGGTT 660  
 GGAGAAAATA TTGATAAATC AAAAACTAAT TTTTAAACT ATGCGGACAA TTCAACTAAT 720  
 ACATCAGATG GAAAGTTTCA TGGGCAACGT TTAAATGATG TCATCTTAAC TTATGTTGCT 780  
 TCAACTGGTA AAATGAGAGC AGAATATGCT GGTAAACTT GGGAGACTTC AATAACAGAT 840  
 TTAGGTTTAT CTAAAAATCA GGCATATAAT TTCTTAATTA CATCTAGTCA AAGATGGGGC 900  
 CTTAATCAAG GGATAAATGC AAATGGCTGG ATGAGAACTG ACTTGAAAGG TTCAGAGTTT 960  
 ACTTTTACAC CAGAAGCGCC AAAACAATA ACAGAATTAG AAAAAAAGT TGAAGAGATT 1020  
 CCATTCAAGA AAGAACGTAA ATTTAATCCG GATTTAGCAC CAGGGACAGA AAAAGTAACA 1080  
 AGAGAAGGAC AAAAAGGTGA GAAGACAATA ACGACACCAA CACTAAAAAA TCCATTAACT 1140  
 GGAGWAATTA TTAGTAAAGG TGAAYCgAAA GAAGAAATCA CAAAAGATCC GATTAATGAA 1200  
 TTAACAGAAT ACGGACCAGA AACGATAACA CCAGGTCATC GAGACGAATT TGATCCGAAG 1260  
 TTACCAACAG GAGAGAAAGA GGAAGTTCCA GGTAAACCAG GAATTAAGAA TCCAGAAACA 1320  
 GGAGAYGTAG TTAGACCACC GGTGATAGC GTAACAAAAT ATGGACCTGT AAAAGGAGAC 1380  
 TCGATTGTAG AAAAAGAAGA rATTCCATTC rAGAAAGAAC GTAAATTTAA TCCTGATTTA 1440

CCAACACTAA AAAATCCATT AACTGGAGAA ATTATTAGTA AAGGTGAATC GAAAGAAGAA 1560  
 ATCACAAAAG ATCCGATTAA TGAATTAACA GAATACGGAC CAGAAACGAT AACACCAGGT 1620  
 5 CATCGAGACG AATTTGATCC GAAGTTACCA ACAGGAGAGA AAGAGGAAGT TCCAGGTAAA 1680  
 CCAGGAATTa AGAATCCAGA AACAGGAGAT GTAGTTAGAC CACCGGTCGA TaGCGTAACA 1740  
 AAATATGGGa CCTGTaAAAG GAGACTCgAT TgTaGGAAAA AGarGAATTc CaTTCaGGAA 1800  
 10 AGAcGTaATT TaTCCTGTTT AGCACCCGGG GCAGAAAAGT TAC 1843

## (2) INFORMATION FOR SEQ ID NO: 278:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 8536 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

TAAACAGCGC GTGTACTTGT GATTCCCCCT TCTTCTATTT TACCCACCCG GGAAATAATA 60  
 25 CTTTTCGCGA TTCCTTACTT GaACAAGCAA TATTTTATCa GCTGTTTCTT CAACTAAACA 120  
 GACACATTTA ATCATCTTTG ACACCCCAAC TTTGTGAAAT CAATTTTTCA AATTATACTG 180  
 TACAATTATG TTATCATATA TGAGTAGTTA TAGCGCAAAA CGTTAGCAAT TCAGCGCACC 240  
 30 CAACTTTTCA TATAAACAGA AGATACTAGG GGGAATTATT ATTATGGCAA AACgTTCCAA 300  
 ATCACAACGT TTATCAAGTT TACTAAATGT CGCAGGTTTC ATAGTCGACG GCTACAATGG 360  
 CTATAAATAT CATGCTAAAA ATAAAAAATT AGTATATCTT TCATTAGGTT TAAGCACTGT 420  
 35 AGGAACCGTG TTAGACTTTT ACATTTCAAT TAAGTCACCA CGTAAGTTCA AAAAAGCAGT 480  
 GGCAGTTGTT ACTTTAATAA CAAACGGTGC TAGATTATTT ACAAGCATTc GCAAAGTAAA 540  
 40 ACATGAATAC TAATTCAGAA AAGGATTGGT CGAACATAGA ACATGAAGTT CATTGACCA 600  
 ATCCTTTTTC ATATACAAAA ATTCCCTATTA CAACAATCAC GACTTTGATA GCCCCGCCAT 660  
 AAAATTTAGG ATTCAATCCA ACTTTTCAGC TTGTGaAATG TAATAGGaAT TCATTATATA 720  
 45 TTTATATACG TAAGACTTTA GTGAATATAT CTATAATTAT TTACTTGGTA AGCTGGTACC 780  
 GTTCTGTAAG TTAAAATACC TGGTGCAGCT GAATAGTTCA TTTCTGAAAC TAAAATACTA 840  
 CCATCGTTAT TTACACGTTc TACAAACATA ACATGACCAT AGTAACCTAC ATCTGTTTGA 900  
 50 GCGATAGAAC CTACAGTAGG TCTATTGTCG ATAGTGTAAC CATCTGCTGC CGCTGCGTTA 960  
 TCCCAGTTAT TAGCATTCCA CCAATAAGTA CTAATACCTT TACCAATTTc AGCACGACGA 1020

55

	GTATTGTAAC CTCTATTGT CGTTGTTGCA GATCCTGAGT TCGTAGATGC aTTACCAGTT	1140
	ACTTTCAATT TTTGACCCGG ATATATAAAG AAATTATTTA AGCCATTTAA GCTCATAATT	1200
5	TTTTGATAAG TTGTACCATA TTTTGATGCA ATTAATGACA ATGAGTCACC TGCTTGACT	1260
	GTATAGTATG ATCCGCCACC TGAGTTCGTT GATGGACGGC TACTATTGCT CGCAGCGTTA	1320
	CTTGAGCTAG CAGTACCTGA TACTTTTAAT TTTTGACCTG GATAAATAAA GAAATTATTT	1380
10	AAACCAITAA GTCGCATAAT GTTTTGGTAA GTTGTAACCAT ATTTTGATGC GATTAATGAT	1440
	AATGAGTCGC CTGCTTGTA TGTGTAGTAT GATCCGCCAC CTGAGTTCGT TGATGGACGG	1500
	CTACTATTAC TCGTAGAATT ACTTGAGCCA GATACTTTTA GTACTTGGTT TGGGAAAATT	1560
15	AGATTAGATG TTAAATTGTT TAATGACTTT AATTTAGCAA TCGAAATCCC ATACTTATTT	1620
	GAAATTGCCC AACTGATT CACCGGTTTT ACTGTGTGAG TTGTAGCCGC ATTTGCTTGA	1680
20	GTTGCCGCAA CAGCGCTAAT CGCGCTTGTC CCAATAATAG CTGCAATTAC TTTTTTTTGC	1740
	ACTTTAAAAT CCTCCTCTTG CTTAACTTTC CTAACATTCT TTTATCCGAA TTTATGAATA	1800
	CTACATCATT ATACGATTTT ATTATGTATA ATAGGTTGAT GTTTGATGAC ATTATGaTTA	1860
25	CAAAAAATC ATATACTGTA TCATCAAAT TTATAATTAT CCCTTAAAAT TATTACAACT	1920
	TATTAGATTT TACAATATCT AAATTATTAC AATTTCATAA TATTTCACTA TAAAATGATT	1980
	ACAATCCCTT TCTCTATTGG AAATAATTTT ATTCTCCAAC AATAACGCCC TACAAACATA	2040
30	AGCATGAACT TTGCTTGTA GACGCAATAT AATTTATTTT GCTAATGACG TTTCTATTGC	2100
	CTTAATCTCA TCTTTAGATA AATTAAACAGG TTTCTCTCCA TCTTTGACAT CTTCTGCAA	2160
	CGCTTTTTGA GCTTCTTTTG AATGATACAA TTCTACGATT TTAGCATATG TTTTGTTATC	2220
35	CAAGTCTTTG TCATTAAGTG CAACAATATT AATATATGGC TTTACTGCAT CTGAATTTGA	2280
	TTTTTCTAAA AATATCGGAT CATTTTTAGG ATCTTTACCC GCTTTAGTTG CTACACCGTT	2340
40	ATTAATAACT GCAATATCGA CATCAGATAA AGCACGTGCA GTTTGTTGTG CATCTACTGC	2400
	AGTAATTTTT AAATGTTTTG GATTTGACGT TATATCTTTC ACCGTGCCTG CTAATCCGAA	2460
	ATCTTTTTTC AGTTTTATTA AACCAGCTGC TTCTAATAGT TTAAGTGACAC GTGCTTGTT	2520
45	TGACACATCA TTTGGAATGA CAACTTTAGC ACCATCTTTA ACCTTTTTGA CATCTTTAAT	2580
	TTTATCTGAG TAAATGCCCA ACGGTGCTAA AACTGTTGTA CTTAATGCTG AAATCTTTGT	2640
	TCCTTTATGC GCCTTTTTAT ATTGATCTAA AAATGCAAAA TGTTGGAATG CATTCAATC	2700
50	AATATCACCA TCATTTAATG CTTTATTCGG TAAATTGTAA TCTGAGAAGT GCTTAATCTC	2760
	CACATCAATA TCATCTTTTT TAGCTAATTC TTAAACCTTC TCCAAGCCT TAGTGTCATT	2820
55		



	AAGTGCTACG	ATAACTAACC	CAATCAATCT	TTTCATTCTA	TCAATTCCTT	TCAAAATCTT	2940
	CACTATATAT	CATTAATGTC	TACGTATGAA	TCTAGCTAGA	ACATTCCCTA	GCGTTTGAAT	3000
5	CACTTGGACA	ATAATGACTA	ATACAATAAC	GGTAATAATA	ATGACCGTCG	TATCAAATCT	3060
	TTGATAACCA	TACACTAAAG	CTAAGTCTCC	TATACCACCA	CCGCCAACAG	CTCCTGCCAT	3120
	CGCCGTACTT	CCAATAAGTC	CAATAATCGC	AGTGGAATT	GCTAATACTA	ACGAACCTAA	3180
10	AGCTTCAGGA	ATTAAAAAAT	ATCTAATGAT	TTGTAGTGGT	GAAGCGCCCA	TCGnTTTCGC	3240
	CGCTTCAATA	ATCCCCTCGT	CTACTTCCAA	TAATGAGTTT	TCAACAAGTC	TTGCAATGTA	3300
	AGGTGCCACA	TATACTGTTA	AAGGCACGAT	GGCAGCAGTC	GTACCAATTG	AAGTACCTAC	3360
15	TACTAATTTT	GTGAATGGCA	CAATCGCAAT	TAACAAAATA	ATAAATGGTA	GTGACCTTAA	3420
	AATATTGATT	AAAGGATTTA	AACTTGATG	TATCACTATA	TTGGGCCATA	TGCCTTGTTT	3480
20	TCGAGTAATT	ACCAATAAGA	CACCTAATGG	AATACCAATC	ACTGCTCCTA	AAAATAAAGC	3540
	AATAGATACC	ATATATAGCG	TTTCGTACAA	TGCTTGTAAT	AACTGTGCAC	TGTCTAAATC	3600
	AGAACCAAAC	ATATGTTAAT	GcACCTCCTC	AAATTGAATA	TTTTTCTCTT	TGAAATATTG	3660
25	ATTTATTGCC	GTGTCTTCAA	ATTGTTGATC	CATATTAAAT	CGAAGCCACA	TATAACATAC	3720
	GGTGTTACCT	TGTATTTCTG	ACATAGATGA	AAATAAAATT	TTAACCTCTC	TGCCACAAAT	3780
	TTGAATCAAG	TCATTTATAA	TCGTTTGTGT	CACCTGAGTT	TCCTCGACGA	AGATTTTATA	3840
30	ATCTTTAAAA	TCGCCAACTT	GTTCGTCATT	CAATCGACGA	ATCAATGATG	TACTTGCTC	3900
	AGTCTGTATA	ACTGTAGACA	CAAAATTTTG	AGCAATCGTC	GTTTTAGGAT	GAATAAACAC	3960
	CTCTTTAAcA	GTTCTGTTT	CAACCACTTT	CCCCTTTTCC	ATTACAGCAA	CACGATTACA	4020
35	AATGTCCTTA	ATAACGCGCA	TTTCATGTGT	AATCATCATA	ATTGTAATGC	CAAAGGTTTG	4080
	ATTGACATTC	TTTAATAACG	TCAATATCGA	AGCAGTCGTT	GCTGGATCCA	ATGCGCTTGT	4140
40	TGCTTCATCG	CATAGGAGTA	TTTTCGGATT	AGTAACAAGC	GCTCTTGCAA	TAGCCACCCT	4200
	TTGCTTCTGC	CCACCAGATA	ATTCATCAGG	AAATTGGTCT	TTTTTATCAC	TCAATCCTAC	4260
	AAATTCAAGC	ATTTCCGTTA	CTCGTTGCTT	AATTTCTGTT	TTGCTTTTCT	TACTTAAAAT	4320
45	GAGTGGCATT	GCTACATTTT	TAAATACGGT	AGCTGAATTT	AATAAATTGA	AATGCTGAAA	4380
	TATCATACCG	ATATCTTTCT	TAATATCCCT	CATCATTTTA	TCGCTATAAT	TCGTAATATC	4440
	ATGTCCATCT	ACAATCACTT	GTCCATTCTGA	GGCAGCTTCA	AGATGATTCA	CGAGTCTTAC	4500
50	CAACGTACTT	TTTCCTGCAC	CACTATATCC	AATCACACCA	AAAATATCAT	TGCGATTGAC	4560
	CGTAAATGAT	ACGTCCTTCA	AAGCATCTAT	TTTTTGCTTC	TTTTTATTAA	AGGTCTTACT	4620

55

	TAAGAGCATT ATATGTAAAA TTGCATATAT CGTCAATACA ATTTGCCGAA TTTTCTAAAA	4740
	AATTAAAAAA TAAGTAATTC ATGTGACAAT GACGAATTGT GAGACTACTA TGACATTTAT	4800
5	CAAATTAAAT CCATAAAAAT GTCCACCAAT CCTCCACAAC GCAATTACTA AATATTAACA	4860
	TCGCACAAAA AAGCACTAGC ATATTCAAGA ACAACAAACG TTGAACCTAA AATATATGCC	4920
	AGTGCTGCTA TTATTTATAA AGTATCTAGT GCTTGTTTTA AATCATCGAC TAAATCTTCA	4980
10	GTATCTTCAA TACCTACAGA AATTCTTACA AGTCCGCTCG TAATACCTTC TTTAGCTCGA	5040
	ATATCTGCTG GAATGGATGC ATGTGTCATC AATGCAGGTA CTGAAATTAA ACTTTCCACT	5100
15	GCACCTAAAC TTTCAGCTAA TGTGTAATAC GATGTTGCTT TAATCAATTG TTTGGCACTT	5160
	TCTGTATTTT TCACTTCAAA TGCAATCACA CCTGTATGGC CATCCGCTTG AGCCATATGG	5220
	ACATCATGAT TTAAATGACT TTCAATACTT GGATGGAACA CTTGTTGCAC AGCTGGATGT	5280
20	GCTTGTAACA TTTTAATAAT TTCAATAACG CTGCGATTAA TTTGTTCCAT ACGTAAACCT	5340
	AATGTTTTAA TACCCCTCAC AAGTAAATAG CTATCTTGAG GTCCTAAAAT GCCACCTGTT	5400
	GAATTTGAAA TAAATGCTAA ACGTTCTGCA AGCTTGTCAT CCGATGTTGC AACTAAACCA	5460
25	GCAACGACAT CACTATGTCC ACCTAAATAT TTCGTTGCAG AATGTAAGAC AATATCGATA	5520
	CCTAAATCTA ATGGATTCTG ATAATAAGGT GTCATAAATG TGTTATCAAC AACTGAAATC	5580
	AAACCGTGTT CTTTCGCAAT TTCAGCAGAC TTTTAAATGT CAGTAACACG TAATAATGGA	5640
30	TTAGAAGGTG TTTCAATAAA CAACATCTTT GTTGTGGGGC GTATCGCTTG TACAATTGAA	5700
	TCTGTATGCG TTGTATCTAC AAAATCCACT TCAATGCCAA ATCGTGTAAT TACTTTTGTC	5760
35	AATGCGCGAT AAGTACCGCC GTATACATCT GAATTTAAAA TAATATGATC TCCTTTGTCC	5820
	AACAGCATAA CAACTGCACT GATTGCTGCA ACACCTGAAC TAAATGCAAA GCCATGTTTG	5880
	CCATTTTCTA ATGTCGCAAT AACGCTTTCT ACAGAACTTC TTGTTGGATT CGCAGTACGA	5940
40	GAATATTCAT ATCCTTGACG TAAATCACCA ATATCATCTT GTAAATATGT ACTTGTTTGA	6000
	TAAATTGGTG TTGTAACGGC ACCTGTATAA TCGTCTGTTG TGTGCCCACC ATGAATTAAT	6060
	TTAGTTTTCT TGTTCAATTAT TATTCTCCTC ATAATTAAAT ATTTGCTTAG ACATATATCG	6120
45	ATCACTACCA TCTGGAAATA CGACAACAAT CGTACCTTCA GATAATTGCG CTTTTAAATT	6180
	CAATGCACCT TGTAATGCTG CACCTGAGA ACTGCCTACT AACCAACCCTT CATTTATAGC	6240
	CAAACCTTTG ACATTTGAA AGGCATCTTG ATCTTTAATC GTAAATATCC CATCTACAAG	6300
50	ACGTCTCTCT AAAAATATCG GCCATTTCTC AGAACCGATA CCTTCAGTGT CATGTGCATG	6360
	AGCTGGCCCT CCATTTAACA CGGACCCTTC TGGCTCAACG GCATAACATT GCACGTGATG	6420
55		

EP 0 786 519 A2

	ATAATCAATT TGCTGTAATG CTGAAGTCAA TTCGGGTCCC AATGTATGAA AATATGTATC	6540
	CGGATTATGT TCGGATTCAA ATTGATTCAT ATAAACGGCA CCATATTTTT CAGCATAGGA	6600
5	ACGTGCAGCT AATTGTGCCC CATGCATACC TTCAGACTGA CTCGTCCTTG AACTTCTGCT	6660
	ACCAAGCGCT ATCATAATAT TAATCTTTTC TTCTGAAAAA CCATACGGCG CAAAGATCTT	6720
	ACATTTCAAA TGATGTCTAT TCGCTGCAAT AGCTAACCCCT ATGCCTGTAT TACCAGCAGT	6780
10	CGCTTCAACA ATAGTTTGAC CTGCACGCAC ACGCCCTTCT TGAATTGCCT TCTCTACTAA	6840
	ATATTTCCCG AGTCTGTCTT TAACACTGCC TCCAGGATTC CATGTGTTCAA GCTTGGCATA	6900
	AATTTTAACT TTATCATCAC TATAATGTTT TAACAGTACT AATGGTGTAT TGCCaATTAA	6960
15	ATCATAAGTA ATCATAGATG CACCCTCATC TGACATGCCG ATCAAATGAA TGAAACCTTT	7020
	CTTCATGTCT CAATTTTAAT TCTTACTTTT CAGATAAGAA TTATAAACGA CATTTTGTTA	7080
20	TTTTGCAATT ATCTAAGTTT CGATTAATTC AGAACCAGTA CTAAATTTTC AATTCCAAAC	7140
	AAAAAACAC CTGAGCAACA CAAATACTTG TGTGTCAGAT GCTTCTATAT ATTAATAAA	7200
	TAATTGCACG ATAAAGACTA AAATAATAAC GACAGGCATC GCATACTTAA TTAAGTAATA	7260
25	CCAACCACTG AATAATCTAA ATCGATCTTT ACCAAAATAT TGTGTGAATA ATTTTTTATC	7320
	TAATAATTGT CCTACGACAA GCGTAGTACC TAATGCGCCT AATGGCATCA ATACATTGCA	7380
	AACGATGAAA TCCATATTAT CAAAAATCGT TCCCGCACCG AATCTTACAT CTTTAAAGAT	7440
30	ACCAAAAGAT AAGGTTGCTG GAATACTAAT GATAAATACT AAAATACTAC CGATCACTGC	7500
	GACTTTTTTA CGTTTTGTAT TGTCAATTCTT CGTGAAGTTA GAAACATTTA ATTCTAATAA	7560
	AGAAATAGAT GACGTTAAAG CCGCAAATAA GAACAGCACT AAGAATCCCA AATAGAATAA	7620
35	TGTGCCTAGA TGCATTTGAC TAAAGACCAT TGGCAGTACT TTAAATAATA ATCCAGGCCC	7680
	TTCTGTGGT TCATAGCCAA AACTATGTAA AGCCGGAAT ATAGCTAGAC CTGCCAATAC	7740
40	AGATACAAAG ATATTCATAA CAACGATAGA AATAGCTGAT GACTTAATCG TCATGTCTTT	7800
	AGAGGCATAA CTCGCATAAG TAATCATACC TGTAGTTCCT AATGATAACG TAAAGAATGA	7860
	TTGACCTAGC GCAAACAAGA TGCCATCAGC AGTAATCTCT GATACTCTTG GTTGTAATAA	7920
45	AAATTTCAAC CCTTCTAAGA CGCCATCTAA TGTAAAGAC TTAATCACAA TGACGATTAA	7980
	AAAGACAAAC AGCAATGGCA TCATAACTTT CGATGCCTTT TCTAATCCTT TTTCAACACC	8040
	TAACATGACA ATAATCATCG TAGCGAATAT GAATATACCT TGCCCTAGAA CGGTAAACCA	8100
50	AGGATTTGAT ATTACCGCTT CAAAATTCAT TTCTTGGAGA TGATTGATGC GTTGAAATAT	8160
	AACTAATTGC CATAATACTT GTCCGATGTA AATGACAATC CAACCACCGA TAACACTATA	8220

55

TTTACCAGTT AATTTACTAT ATATTTGTGT TGTATATGTC CGTCCCATT TCCCAACAGT 8340  
 GAATTCCATA ATGAGTAATG GCAACCCAAC AAAAATGGTG AATATTAAGA ACATAGCTAG 8400  
 5 AAAGGCACCG CCGCCATAAA TCCCTGCCAT ATATGGGAAT TTCCACATGG CACCAAGACC 8460  
 GATTGCAGAA CcCGCACTAG CTAAAATAAA TCCAGTTGAT GACTTCCATT GTGATTGTTG 8520  
 TCTTTTCATC ATTCAC 8536

(2) INFORMATION FOR SEQ ID NO: 279:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4328 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

GCTTTGGCCA TTTTATGTGG CGATTGAGAC AATCtGtKGT TGTCTTATTT GATGTTGTAT 60  
 TTCAACTGGT AATTCTAGTT GCGATTGAAA TAATGGCAAC TTTTCCCAAT CATTAAACAAA 120  
 25 TAATTCAATA CCGCTATGT CTAATACTTT AGCACGTGCA TCATCAACAA GACGGCGTTC 180  
 CAATTGATTT GCTTCTTCTT TAATACCTGG TGACGTACTT TCTAATATCA AATTAGATAT 240  
 AGGGATGTGA CCATTAATTG CATAATATAA TGCAACACGC CCACCCATTG AATATCCAAA 300  
 CAATGTTATT GATTTATCTT TATATTTATC TAAAATTTCGG TCTAACAACG TCGTAATATA 360  
 ATCAAAATTC CACGTTTCAT CCATTGAAGA CTGATCTTCG CCATGGCCTG GTAAGTCTAT 420  
 35 AGTGATGACA TGATAGTTAT CAGTAAATTT TTCGATGTGA TTATGATAAG TACGGCTGTC 480  
 GCTAAGAAAT CCATGCAGAA ATACTAAAAC TTGATTGGTC TCAACGTTTG CTTCATAAAA 540  
 TTTATAATGT GTCATGAATC ATTTCACTCA ATTTCTGGTA TAAAATTTGA TGCTGTTTAA 600  
 40 AGTTATCTTC GCGATTCGTT ATCAATTCAT AAATCGTCGA AGTTTCAGAT AACAAATGTGG 660  
 CATTTTTAAA TTCTGAAACA CTGTTAAAAC GTTTAAAATC GAATTGATAT AACTTAGCTG 720  
 TATACTCGAA ATCCAATCCC GTCGGTGTGC CAAACAACCG TTCAAAATAG TCAGTTGCAC 780  
 45 TTTCTTTTTG TGGTAAATAT GAAAAAATAC CGCCACCATC GTTGTTCAAT AATACAATAT 840  
 TCATCTGAAT ATTATTTAAT TTTGACATTA ATAGTCCATT CATATCATGA TAAAATGATA 900  
 AATCACCTAT CAATAATGTT ATTCGTTTAT GCACAGCCAT ACCCAGTGCA GTTGAAACGA 960  
 50 TACCATCAAT ACCATTCGCA CCACGATTCG CATAGACATC TATATTTTTTA TTCAATAACA 1020  
 AGTTATCTAC ATCTCTGATA GGCATACTAT TACTAATAAA TAATGCATCT TTTTCAGATG 1080

	TAATTTCTTT	ACGCCCTTTT	TTCTCTAAGC	ATTGCCATTT	TTCTAACCAA	CTTACGCGAT	1200
	TAACTGTCGT	GTCTTCCATT	AATGACCTAA	AGAAATCATT	CGCAGAAATC	TCATATGAAA	1260
5	TATCTGGCGC	TATCGGAAAG	ACATCAATCT	TATCATTGTT	TTGCACTAAT	ATTTGAAATG	1320
	CATCAGTTTT	CTTTAACCAT	TGATTTAACT	TTTTAGAAAT	CACTGGTTTC	CCAACACGAA	1380
	TTACGAAATC	CACATTTAAG	TCTAAGCCGC	TTCTAAACAG	CAAATCATAT	GTACAGATAA	1440
10	CATTCGGATG	ATCAAATTTT	CTTAAATGAC	TTAAAGGATC	AGCTAAAATA	GGCAAATCAT	1500
	ATATCGTTGA	ATACGTTAGT	ATTTGaTCAA	CTTCTTGGTG	CTGCATATCC	CCTACAATAA	1560
15	TTAAACCTTT	TTTCTTATTT	AAAATGTGTC	TTAATGCCGA	TGCATCTATA	CTTTTTTGAT	1620
	AGTGCGGTAA	AATCTTCATC	TCAGAAGTTA	ACAATTCTGT	TGCATTCAAA	TCAGGTGTTA	1680
	ACGGATCTCT	AAATGGCAAG	TTAAAATGAA	TTGGCCCTTT	ATGTGGTCCA	TATAAATATT	1740
20	GACTAGCAAT	TTGCATTTGA	TAGTAAATTG	CATCAATGGT	CTCTTTACTA	TCATCCGCAA	1800
	TAGGCATATC	GAATCATAA	CTTACATAAT	TATTAAACAT	ATTTACTTGA	TTAATCGCTT	1860
	GTGGTGCGCC	TACACTTCTT	AATTCATGCG	GACGGTCACT	TGTTAAAACG	ATTAAAGGAA	1920
25	TTCTACTAAT	TTGGCTTTCA	GCAATTGCAG	GCGTATAATT	CGCTGCTGCT	GTACCTGACG	1980
	TACATAATAT	AGCGACAGGT	CTTTCACTGC	CTTTAATTAA	CCCAACTGCA	AAAAACGCTG	2040
30	CACTTCGCTC	ATCGGGGTGT	ATCCATGTTT	TAATATTTGG	ATGTGCTTCA	AATGCAAGTG	2100
	CAAGTGGCGT	TGAGCGTGAT	CCCGGACTGA	TAATACTTTC	CCTTACGCCG	TACGCATATA	2160
	ACTCAGATGC	AAATGTAAAA	ACTTGCTTCG	TTAAAGCTGC	TTTATGATTT	CCCATTCTATA	2220
35	TCGACTCCTA	ATGCATTCAT	CATAGGTGTG	AACTTAAGGT	TCGTTTCTGC	CAATTCACTA	2280
	TCTGGATCAG	AATCTTTAAC	AATGCCACAC	CCAGCAAATA	AAGTTGCTTG	TGCTTTCTTA	2340
	ATAAGCATCG	AACGAATTGC	AACAATAAAT	TCACAATCAT	CGTATATATC	TATATAGCCA	2400
40	ACCGGTGCAC	CATATAATCC	TCGCGTACCA	AATTCCTTCT	GCTCAATAAA	ATCCATTGCA	2460
	AATTCCTTTG	GATAGCCACC	TAAAGCAGGT	GTTGGATGTA	AATTATCAAT	TAAACTAATA	2520
	TACGAATCAT	CCTTCAGTGG	CGCCTTTATT	TCAGTGTACA	AGTGATATAA	ATGATCATTT	2580
45	TTTAGAATTT	TAGGCGTCTT	ATCATAATGT	AATTCAGTGA	TATAAGGTTT	AATATCATGT	2640
	AAAATACTGT	CAACAACAAA	TCGATGTTCT	ATTAAGTTTT	TATTATCTTT	TAAAAATGCT	2700
50	TCAACATTTT	TTGTATCTTC	GTCCTCATCT	TGTGAACGTT	TAATTGTACC	TGCTACAGCT	2760
	TTAGTCGATA	GTATTTTATT	ATTGACCTTT	ATTAATTGTT	CAGGTGTTTG	TGAAAAGAAT	2820
	ATAGAATCTT	GTGATTCTAA	CAAGAATATA	TAACIGTTTT	TTTCTTTAGA	ATATGCTTGC	2880
55							

ACAATTTTTT CTTCATTATT AATAGATTCT ATAGCTTCTA CTACAAGTTG ACGCCAGTCA 3000  
 TCTTTATAAA TATCTTCATT TCTAGTAATT TCCCCAATTT GCTCGTCCAC ATCTATGTCC 3060  
 5 GATATATTGT TGAACAAATC CATTAAATCG TTCAATGCCT CAACAGTAAA ACTTTCCTTT 3120  
 TTAAGTGTAT AAGTTAAAAA TGTCCCATTA TTATCAGTTG AAATTAAAAAC TTCAGGTAAT 3180  
 ACAAAATGAT TTAGTCCAAA CTCTCGCCAT TCATCATCTG ATTTATGACT TGAAAATTGG 3240  
 10 AACCTCCAA CAACTCGAAG ATGATGTTTC TCAGATTGCG GATGTATAAA TGTGATGTTA 3300  
 TGTTTTAAAT TTTCCAGTC TTTAAAAATA GATTGTTTAT TTTTAGAATT ATTTTGAAT 3360  
 AATTGAATTG CTTTGTAGCC AAAATATGAC GTTCGATTAT CATTCAAACG CATATAAAG 3420  
 15 CGATCTCCTG CCTCATTGTC AGTGAGATGA AATAATGTGC TCGGGTCTAG TGAAGTGTAT 3480  
 AATTTCACTT CAACTGAAAC CCATTCTTTT GAGCTGCCAT ATATCTCTTT GACAATATCG 3540  
 20 TCCTCTAATA CGCCCGTAGC CATCCATTTC ACTTCTTTCT TCGTCTTTT TCACTCATT 3600  
 TTATATTGTA TCATTTTGGG ATAATTGTGT TACAAGAATT GCTTAACTT ATCTTGCAAT 3660  
 TTTTCACGTC AATTGACCTT TATGCTACTT TCTATTAAAA TATCTTTGTT ATAAAAAATA 3720  
 25 TGATTTAAAG AGGTTTTGTA TTCAATGAGT AATCAATATC AGCAATATTC TACAGTTAAG 3780  
 AAATATTGGC ATTTAATGCG TCCTCATACA TTAAGTGCTT CCGTAGTACC CGTTTGTAGT 3840  
 GGTACAGCAG CATCTAAAAT ATATTTCTT GGTAGCGAAG ATCATATTAA AATCAGCCTA 3900  
 30 TTCATTGCCA TGTTACTAGC ATGCTTACTT ATTCAAGCAG CAACTAATAT GTTTAATGAA 3960  
 TACTATGATT ATAAAAAGG CCTCGATGAT CATGAATCTG TAGGCATTGG TGGTGCCATT 4020  
 GTTCGCAACG GTATGAGCCC AGAGCTTGTG CTACGATTAG CCATTGCATT TTACATCTTA 4080  
 35 GCAGCAATAT TAGGTTTGTT TTTAGCTGCT AACTCTTCAT TTTGGTTATT ACCAGTTGGA 4140  
 TTAGTATGTA TGGCTGTTGG TTACCTATAT ACAGGTGGCC CTTTCCCTAT TTCATGGACG 4200  
 40 CCTTTCGGTG AATTATTCTC AGGCGTATTT ATGGGTATGT TTATTATCGT TATTGCATTC 4260  
 TTTATTCAAA CTGGCAATAT TCAAAGTTAT GTAATTGGT TAAGTGTACC TATAGTAATC 4320  
 ACTATCGG 4328

45 (2) INFORMATION FOR SEQ ID NO: 280:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1450 base pairs  
 (B) TYPE: nucleic acid  
 50 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

	GTTCAATACA GAAAAAATAA ATTTAGATGT TGAAGCATCC TACAATTAAT ACAGATCCAT	60
	TTCAATATAT TTAAACTAAA ATCTCGGGAT TTCTAAATTT TGAAATTTTCG AGGTTTTnAT	120
5	ATTTTTATTT AAAATAGCAC ATTTATACTT TATAATAGTA AAGATGAACA TATAAGGAGG	180
	CCAAATCATG GCAAAACATC CATTGCAACA ATTTAATCTA GAATCTAGTT TAATTGACGC	240
	TGTGAAAGAC CTTAATTTTG AAAAACCAAC TGAAATTCAG AATCGAATTA TTCCAAGAAT	300
10	ACTAAAGAGA ACAAATTTAA TTGGTCAATC TCAAACGGGT ACAGGGAAAT CTCATGCATT	360
	TTTATTACCA TTAATGCAGT TAATTGATAG TGAAATAAAA GAACCACAAG CAATCGTAGT	420
15	TGCACCAACA AGAGAACTTG CACAACAACT ATACGATGCA GCGAACCATT TAAGCCAATT	480
	TAAAGCTGGT GTTTCAGTTA AAGTTTTTAT TGGTGGTACA GATATAGAGA AAGATAGACA	540
	ACGTGTGAAT GCACAACCAC AATTGATTAT AGGCACCCCT ACTAGAATTA ATGACTTAGC	600
20	TAAAACGGGA CATTTACATG TGCACTTAGC ATCATATTTA GTTATTGATG AAGCGGATCT	660
	TATGATTGAC TTAGGATTAA TTGAAGATGT AGATTACATT GCTGCAAGAT TGGAAGATAA	720
	TGCAAATATT GCGGTGTTTA GTGCTACAAT CCCACAACAG TTACAACCAT TTTTAAATAA	780
25	ATATTTAAGT CATCCAGAAT ATGTAGCTGT CGACAGTAAA AAACAAAATA AAAAGAACAT	840
	CGAATTCTAT TTAATACCTA CTAAAGGTGC AGCTAAAGTT GAAAAGACTT TAAATTTAAT	900
	TGATATACTA AATCCATACT TATGTATTAT TTTCTGTAAT AGTAGAGATA ATGCAAATGA	960
30	TTTAGCACGT TCACTAAATG AAGCTGGTAT TAAAGTTGGT ATGATTCATG GTGGCTTAAC	1020
	GCCaCGTgAA CGTAAACAAC AAATGarACG TATACGTAAT TTaGAATTCC aATACGTTAT	1080
35	TGCCaGCGAT TTAGCATCTC GTGGTATTGA TATTGAAGGT GTTAGTCrTG TCATCaATTT	1140
	TGATGTGCCA AATGATATTG ACTTCTTTAC GCATAGAGTC GGACGAACTG GTCGTGGGAA	1200
	TTATrrAGGT GTAGCAATTA CGCTTTATAG TCCTGATGAA GAACACAATA TTTCATTAAT	1260
40	AGAAGATCGC GGTTTTGTAT TCAATACTGT TGATATTAAA GATGGTGAGT TAAAAGAAGT	1320
	TAAAGCGCAC AATCAGCGTC AAGCAAGAAT GCGCAAAGAT GACCATTTAA CTAATCAAGT	1380
	GAAGAACAAA GTTCGAAGTA AAATTAAAAA CAAAGTTAAA CCAGGTTATA AGAAGAAATT	1440
45	TAAACAAGAA	1450

(2) INFORMATION FOR SEQ ID NO: 281:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1139 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

5 AGTCAGGTAT ATCATGCCaT yCTGAATTGG TCGATATTAA TATCAGTGGT GTTAAAGAAC 60  
 GAATTGTATA CCAATAGACG CTTTATATTG TAAAATAGTA TTAAATGCaG AATAGAGAGG 120  
 AGATTTAATG CGATATGACA AATTATAAAG TTGTCGTTTT AGATATGGAT GACACATTGC 180  
 TAAATTCAGA TAATGTGATA TCAGAAGAAA CTGCAAATTA TTAAACAGCA ATTCAAGATG 240  
 10 AAGGTTATTA TGTGTCTTA GCATCTGGTA GACCTACTGA AGGTATGATT CCAACTGCTA 300  
 GAGATTTAAA ATTACCTGAA CATCATAGCT ATATTATTAG TTATAACGGT AGTAAAACGA 360  
 TTAACATGAC TAATGAAGAA GTAGAAGTAA GTAAATCGAT TGGTAAGCAA GATTTTCGATG 420  
 AAATTGTAGA TTATTGTCGA GATAGAGGCT TTTTCGTTCT TACATATCAT GATGGTCAAA 480  
 TTATTTaCGA CAGCGAACAT GAGTATATGA ATATTGAAGC AGAATTAACA GGTTTACCGA 540  
 20 TGAAACGTGT TGATGATATC AAAGCGTATA TTCAAGGCCA TGTACCCAAG GTCATGGGTG 600  
 TAGATTATGT AGCGAATATT ACAGAAGCTA GAATTGATTT GAATGGTGTG TTCAATGATA 660  
 ATGTAGATGC TACGACAAGT AAGCCATTCT TCTTAGAATT TATGGCCAAA GACGTTTCAA 720  
 25 AAGGTAATGC AATTAAAGCG TTATGTCACA AATTGGGATA TTCGGTGGAT CAAGTCATTG 780  
 CTTTGGTGA TAGTATGAAT GATAAATCAA TGTTTGAAGT CGCAGGTCTA GCTATTGCTA 840  
 TGGGGaATGC ATCAGATGAA CTTAAGCAAT ATGCAAATGA AGTTACGTTG GATCATAATG 900  
 30 AAAATGGTAT TCCACATGCG CTCAAAAAAT TGTTATAAAT TTTAAATAA GCCTTAACAC 960  
 ATGATATTTG AATAAGATAT CTTGTGGTTA AGGCTTTTTT TTTTGTGAA AATGACTTCA 1020  
 GTTATACTAT GGAGGATTTG AAATACATAT TTTAGATTAG TAATGATATC AAACGAATAG 1080  
 35 AGTAAATGTA TATTTtTGTA ATAAATCAAG TATTAAGTAG TCACGGAAGG nAGATAAAT 1139

## (2) INFORMATION FOR SEQ ID NO: 282:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2931 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

50 TCTAAAAATG CTGTGAAATT CTTTATAAAA TATCTAAAAG GAATTAATGT TGATAACATT 60  
 GCTGTGATAG GAAGTAAGAC AGCGCAATAT TGTGAATCAC TTGGCATTCT AGTTGATTTT 120  
 AtGCCAAACG ACTTTTCTCA AGAAGGATTT TTAAATCAT TTAATCAAAC TAACCAAAAA 180

55



	AATGAAGTTG TTAAAATAGA TTTATATACT TCAGTGCCTA ACAAACAAAA TATACAAGAT	300
	GTAAAGAAA TGATAGAACA TCAACAAATC GATGCATTAA CATTTCACAG TTCGTCGGCA	360
5	GTACGTTATT ATTTTAATGA AGGATTTGTA CCAAAATTCA AGTCGTATTT TGCTATTGGA	420
	GAACAAACAG CACGGACCAT TAAATCATAT CAACAACCAG TAACAATTGC AGAAATTCAA	480
	ACACTCGAAT CACTAATTGA AAAGATTTTA GAAAGTAGGG GCTAAAAATG AAATTTGATA	540
10	GACATAGAAG ATTGAGATCA TCAGCGACAA TGAGAGATAT GGTTAGAGAG AATCATGTAA	600
	GAAAAGAAGA TTTAATATAT CCAATTTTGG TAGTTGAAAA AGACGATGTG AAAAAAGAAA	660
15	TTAAGTCATT GCCAGGTGTA TACCAAATCA GTTTGAATTT ACTTGAAAGT GAATTAAAAG	720
	AAGCTTATGA CTTAGGCATA CGTGCCATTA TGTTTTTCGG TGTTCCAAAC TCAAAaGATG	780
	ATATAGGTAC TGGTGCATAC ATTCACGATG GTGTTATTCA ACAGGCAACA CGTATTGCTA	840
20	AAAAAATGTA TGATGACTTA TTAATTGTTG CAGACACTTG TTTATGTGAA TATACTGATC	900
	ATGGTCATTG TGGCGTGATT GATGACCATA CACATGACGT TGACAATGAT AAATCATTGC	960
	CACTACTTGT TAAAACAGCA ATTTCTCAAG TGAAGCTGG TGCTGATATT ATTGCGCCAA	1020
25	GTAATATGAT GGATGGTTTT GTTGCTGAAA TTCGTCGTGG ATTAGATGAA GCCGGCTATT	1080
	ACAATATTCC TATAATGAGT TATGGTGTCA AGTATGCATC AAGTTTCTTT GGACCTTTTA	1140
	GAGATGCAGC AGATTCAGCG CCATCATTGG GGGATAGAAA AACGTATCAG ATGGACCCTG	1200
30	CTAACCGTTT GGAAGCACTT CGTGAATTAG AAAGTGATCT TAAAGAAGGG TGCGACATGA	1260
	TGATTGTTAA ACCTGCTCTA AGTTATTTAG ATATAGTTCTG AGATGTTAAA AATCATACGA	1320
	ATGTTCCAGT TGTTGCATAT AATGTGAGTG GAGAATATAG TATGACTAAA GCAGCGGCAC	1380
35	AAAATGGTTG GATAGATGAA GAACGTGTCG TTATGGAACA AATGGTTTCA ATGAAACGTG	1440
	CAGGTGCTGA TATGATTATT ACGTATTTTG CAAAGGACAT TTGTCGCTAT TTAGATAAAT	1500
40	AAGGTTTTAT ATTTATGATT TTCCATAAAC TGTAGGAGGA ATTTACTTTA TGAGATATAC	1560
	GAAATCAGAA GAAGCAATGa AGGTTGCTGA AACTTTAATG CCTGGTGGTG TAAATAGTCC	1620
	AGTACGCGCA TTAAATCAG TAGATACACC AGCAATTTTT ATGGATCACG GTAAAGGTTT	1680
45	AAAAATTTAT GATATCGATG GTAACGAGTA TATCGACTAT GTACTAAGTT GGGGACCACT	1740
	TATTTTAGGA CATAGAGACC CTCAAGTTAT TAGTCATTTA CATGAAGCAA TTGATAAAGG	1800
	TACAAGTTTT GGTGCATCAA CATTACTTGA AAATAAATTG GCGCAgcTCG TTATTGACCG	1860
50	AGTACCTTCA ATAGAAAAAG TCGTATGGT GTCATCTGGT ACAGAAGCTA CATTGGATAC	1920
	TTTAAGATTA GCACGTGGTT ATACTGGCAG AAATAAAATT GTGAAATTTG AAGGTTGCTA	1980
55		

GCCGGATTCT CCTGGTGTGC CTGAAGGTAT TGCTAAAAAT ACAATTACAG TTCCATACAA 2100  
 TGATTTAGAT GCACTTAAAA TCGCTTTCGA AAAATTGGA AACGATATTG CTGGTGTAAAT 2160  
 5 CGTAGAACCT GTTGCTGGTA ATATGGGTGT CGTACCGCCG ATTGAAGGTT TTTTACAGGG 2220  
 ATTAAGAGAT ATTACGACTG AATACGGCGC ATTGCTAATT TTCGATGAAG TAATGACTGG 2280  
 10 TTTTACAGTC GGTATCATT GTGCACAAGG TTACTTTGGT GTGACACCAG ATTTAACTTG 2340  
 CTTAGGAAAA GTTATCGGTG GAGGACTACC TGTAGGTGCA TTTGGTGGTA AAAAAGAAAT 2400  
 CATGGATCAT ATAGCACCAT TAGGAAATAT TTATCAAGCG GGTACGTTAT CAGGAAATCC 2460  
 15 TCTTGCAATG ACAAGTGGTT ATGAAACGTT AAGCCAATTA ACGCCAGAGA CATATGAGTA 2520  
 TTTTAATATG TTAGGCGATA TACTTGAAGA CGGTTTAAAA CGTGTATTTG CTAAACACAA 2580  
 TGTACCAATA ACTGTAAATA GAGCAGGTTC AATGATTGGT TATTTCTTAA ATGAAGGACC 2640  
 20 TGTAACATA TTTGAACAAG CGAATAAAAG TGATTGAAA TTATTGTCAG AAATGTATCG 2700  
 AGAAATGGCA AAAGAAGGTG TGTTTTTACC ACCATCTCAA TTTGAAGGTA CATTCTTATC 2760  
 TACGGCACAC ACGAAAGAAG ATATTGAAAA ACGATTCAA GCATTGATA CGGCTTTAAG 2820  
 25 TCGTATTGTA AAATAAATAT ACGGACAAAT TGAGAGCCTG AACTTTGTTC AGGCTCTTT 2880  
 TAAATGTATA TAAGGCATGG GCGGCGACTT GATAGTGAAA GTCCACTACT A 2931

30 (2) INFORMATION FOR SEQ ID NO: 283:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1421 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:  
 40 AATTATGAAT GCATTACCAG TATTATTACA AAAGAACAAT TAAAAATGTT TGTTTATGAT 60  
 TATGATACGC ATCTCATTAA AAATGTAaTG GTTGACAGCAG ACGTGTTAAA GGCAAATGAT 120  
 ATTCAAGGAC ATGAACCATT AATCGTTAAC CTTCAAACGA TTGATGAAAC ATTACATCGT 180  
 45 TTACCTATGC ATAATAGAAA AGACATGATG GTTAATGGCG GTGTACTTAT GGCACATTTA 240  
 AATGCCAAAA GTGGTCCGTG GTTAAAAGAT GTGCTAAGAC AAATTGAGAT AGCGATTGTA 300  
 50 ACAGGTAAAG TAAGCAACGA AGAAACTGAA ATTTTGAAAT GGGTGGATAA TCATGTCAAA 360  
 ATATAGTCAA GATGTACTTC AATTACTCTA TAAAAATAAA CCGAATTATA TATCTGGACA 420  
 AAGCATTGCG GAATCACTTA ATATTTACG CACTGCAGTA AAAAAAGTGA TTGACCAATT 480

55

CCCAGATATT TGGTATCAAG GTATAATAGA CCAATATACA AAAAGTTCTG CTTTGTTTGA 600  
 TTTTAGTGAA GTATACGATT CAATAGATT CACACAACTT GCTGCGAAAA AGTCACTTGT 660  
 5 TGGAAATCAA TCTTCATTTT TTATCTTGAG TGATGAACAA ACGAAAGGTC GTGGGCGATT 720  
 TAATAGACAT TGGAGTTCTT CAAAAGGGCA AGGACTTTGG ATGTCTGTCG TGTTAAGACC 780  
 10 TAACGTTGCA TTCTCAATGA TATCTAAATT TAATTTATTT ATTGCATTAG GGATAAGAGA 840  
 TCGGATTCAA CATTTTAGTC AAGATGAAGT CAAAGTGAAG TGGCCGAATG ATATATTTAT 900  
 TGATAATGGT AAAGTGTGTG GTTTCTTAAC TGAAATGGTT GCTAATAATG ATGGTATAGA 960  
 15 AGCAATAATA TGTGGTATAG GTATTAATTT GACGCAACAA CTAGAAAACCT TTGATGAAAG 1020  
 TATTAGACAT AGAGCAACAA GTATACAATT ACATGATAAA AATAAATTAG ATAGATATCA 1080  
 ATTTTLAGAG ATATTACTTC AAGAAATTGA AAAAAGATAT AATCAATTTT TAACGTTACC 1140  
 20 TTTTCTGAA ATTCGTGAAG AATATACTGC AGCTTCTAAT ATTTGGAATA GAACGTTGCT 1200  
 ATTTACAGAA AATGATAAAC AGTTTAAAG ACAAGCAATT GATTAGATT ACGATGGCTA 1260  
 25 TCTAATTGTT AGAGATGAAG CGGGTGAATC ACACCGTTTA ATTAGTGCAG ATATAGATTT 1320  
 TTAACACTAA AGCAAGGAGA GATAGCTATG GGTATGGCAA CCTATGCCGT TGTGGATTTG 1380  
 GAAACAACAG GCAACCAATT AGATTTTGAC GATATCATTC A 1421

30 (2) INFORMATION FOR SEQ ID NO: 284:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2202 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

CCAAGTTGCC TAAAATGATT AAGCAAGGTT TATACCCTAT GCAACGATT GAACAAGAAT 60  
 CTGGAGCCAT CCGACTGCCA ACGATTTCTA GAGTGAGCG TTCATTACAA TGGGGTAATG 120  
 45 ATGCTTATAC AATGATTTTA GATCGTATGA ATATTGAAAC AAATGAATAA TAAATGAACG 180  
 ATAAACAATG GTTATCTATC TGCATAATA AGGTAGATAA TCATTGTTTT TTCACGAAAA 240  
 AATTTACAGA GTAAAAGAAC TTAAATTTCA TATTAAGTCT TTAGAACTCG AACTTAAAA 300  
 50 ATGCTATAAT CATATGTATG TTAAAAAGG AGTTTCGGAA AATGTATGAC ATTAAAAAAT 360  
 GGCGCCATAT TTTTAAATTA GACCCAGCTA AACATATTTT AGATGATGAT TTAGATGCGA 420  
 TTTGTATGTC TCAAACAGAT GCAATTATGA TTGGTGGAAC TGATGACGTT ACTGAAGATA 480

55

CAAACATCGA AAGTGTAAATG CCTGGTTTTG ATTtTTATTT TGTACCTACA GACTGAACA 600  
 GTACAGATGT TGTATTTTAC AATGGTACAT TATTAGAAGC GCTTAAAACA TATGGACATA 660  
 5 GTATAGATTT TGAGGAAGTA ATATTTGAAG GGTATGTCGT GTGCAATGCT GATAGCAAAG 720  
 TGGCAAAACA TACCAAAGCA AATACAGATT TAACAACAGA AGATTTAGAA GCATATGCCC 780  
 10 AAATGGTCAA TCATATGTAT CGATTACCGG TTATGTATAT AGAGTATAGT GGCATTTATG 840  
 GCGACGTATC AAAGGTTCAA GCTGTCTCAG AACATCTAAC AGAAACGCAA CTTTTTTATG 900  
 GTGGCGGTAT TTCCTCAGAA CAACAAGCGA CAGAGATGGC AGCTATTGCA GATACAATTA 960  
 15 TCGTCGGTGA TATTATTTAT AAAGATATTA AAAAAGCTTT AAAACAGTA AAAATAAAGG 1020  
 AGTCTAGTAA ATGAATGCGT TATTAAATCA TATGAATACA GAGCAAAGTG AAGCTGTAAA 1080  
 GACAACAGAA GGACCATTGT TAATTATGGC AGGTGCTGGT TCAGGGAAGA CACGTGTTTT 1140  
 20 AACACATAGA ATTGCTTATT TATTAGACGA AAAAGATGTC TCACCATACA ATGTTTTGGC 1200  
 TATTACTTTT ACAAATAAAG CTGCAAGAGA AATGAAAGAA CGTGTTCAA AATTAGTAGG 1260  
 TGATCAAGCA GAAGTTATTT GGATGTCAAC ATTCCACTCA ATGTGTGTTT GTATTTTACG 1320  
 25 TCGTGATGCA GATCGAATTG GTATAGAACG CAATTTTACG ATAATTGATC CTACAGACCA 1380  
 AAAATCTGTT ATTAAAGACG TCTTAAAAA TGAAAATATT GATAGTAAAA AGTTTGAACC 1440  
 30 TCGTATGTTT ATCGGTGCGA TCAGTAATTT GAAAAATGAA CTTAAACAC CTGCAGATGC 1500  
 TCAAAAAGAA GCCACAGATT aTCACTCgca AwTGGTaGCA ACgGTTTaTA GTgGATATCA 1560  
 ACGCCAATTG TCACGTAATG AAGCGTTAGA TTTTGATGAC CTTATTATGA CAACGATTAA 1620  
 35 CTTATTTGAG CGTGTACCAG AAGTTCTAGA ATATTATCAG AACAAATTCC AATATATTCA 1680  
 TGTAGATGAG TATCAAGATA CTAATAAAGC ACAATACACA TTAGTTAAAT TATTAGCAAG 1740  
 TAAGTTTAAA AACTTATGTG TTGTAGGTGA CTCAGATCAG TCAATTTATG GTTGGCGTGG 1800  
 40 TGCTGATATT CAAAATATCT TATCATTTGA AAAAGACTAT CCAGAAGCGA ATACAATCTT 1860  
 TTTAGAGCAA AATTATCGTT CGACGAAAAC GATTTTAAAT GCGGCTAACG AAGTGTTTAA 1920  
 45 AAATAATTCT GAACGTAAGC CAAAAGGACT GTGGACTGCA AATACGAATG GTGAGAAAAT 1980  
 TCATTACTAT GAAGCAATGA CGGAACtGAT GAAGCGGAAT TTGTAATACG AGAAATTATG 2040  
 AAGCATCAAC GTAATGGTAA GAAATATCAA GATATGGCAA TTTTATATAG AACGAATGCA 2100  
 50 CAATCACGTG TACTTGAGGA AACATTCTATG AAATCTAATA TGCCATACAC AATGGTTGGT 2160  
 GGCCAAAAGT TCTATGACCG TAAAGnAATC CAAAGATTTA TT 2202

(2) INFORMATION FOR SEQ ID NO: 285:

55

(A) LENGTH: 785 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

10 AGTGGTGCAA AGATAGGCAT TGATaATACC GCTAAGCCAG CAAGATGATG GCACGATAAA 60  
 ACCTAAACAG AAGAaATAA ATAGTAATAC GATGATAAAT AATGGTCCAC TCATATGTTG 120  
 AACTAAAGAT GATGAAAAGT GTAAGATTGT ATCTGAAATC ATACCTTCAT TCAACACTAA 180  
 15 ATTAATACCT CGAGCTAAAC CAATAATTAA AGATACACCT ACTAAACTTG ATGCACCATT 240  
 GACAAATGCA TCTACAGTTC CTTTTTCTCC CAATCCAGAT TTACCTGTCC CAGCAATAAA 300  
 CATTATTATA ATTGTAAATA TTAAAAATGC TGAAGCCATA ACTGGGAACC ACCAACCTTG 360  
 20 CGTCATAACT CCCCATACCA TAATTGGAAA TGGTAGTACA AATAATGTAA GGATTATCTT 420  
 CTTACGCAAA GTAAAATGGG CACTATCGTC ATCTTTTAAT ACAGACCATT GCTGTTCAAA 480  
 AGCATCTTTG TCTTCATAAG AATATGACGC TTTAGGATCG TTTTAAATTT TTTTACAGTA 540  
 25 CCAATATAAA TAATAATAA CAAAAATCGC ACCGACAATA CAAGCACCTA TTCTCCAATA 600  
 CAAGCCATCC GTAAAAGTTG TACCAGCGGC ATTAGAGGCA ATTACAACCG AGAACGGGTT 660  
 AATAGTTGAA AATGTACTAC CGACAGAGCT GGcAAGGAAT ATGGCACCAA CTGGAAACGA 720  
 TAGAATCGTA TCCTAACGCT AATAAATATA GGGACTAAAA TCGGATAAAA TGCTACAGCT 780  
 TCTTC 785

35

(2) INFORMATION FOR SEQ ID NO: 286:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 812 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

45

CTAACGnGAT AAGGTTGCAA nTTTATCTGA ACATCTGATG ACTGTAATTT TGTtAATGAT 60  
 AAAATATTTG TCACTAATAG ATATAAATAC TGACTTTCTT GAAAACTATG TACAAGTAAT 120  
 50 TGTTCCTTTT CTATGATAGA CATATCTTTA CTATGTGATA CTAAAATATC TAAATkTCCC 180  
 ATAATTGTTG TTAACGGTGT ACGTATGTCA TCGGAAATTG ATCTTAAAAA ATTTGAATGT 240  
 GTCAGTTGAC GTTCAGCCTG TAACATGGAT TCTCTCGTTT GTTTAAGTAA CGTCACATTT 300

55

ATCACTTGAG AACTTTGGTA ATCAATGGCT AGAATGCCTT TAATCGGAGA TGTGCCAATT 420  
 GGTATCAACC ATTTATTAAT GCCTGGAAAT GTATCTGTTG TTGCACCAGC TTGTCTTTCA 480  
 5 TTTTAAATTA CCCAGCTTAA TGCTTGTTCa TGCTGTTGAG TCGTATTATC GATATGGTTT 540  
 TGCAATGGTA TTGTTTTAAT TACTTTCGAT TGATTGATAA CGTATATAGT AATTGATTGT 600  
 TGCAATAATT GATTAAATTG GTATCCAGCA TTTATTAGTA AGTTTCAAC TGTATAAGTT 660  
 10 TGTTTAATCG AATCATTAAA TTGAAATAAT AAATCTGTAC GATAAAGTTG CTTTTTAGTA 720  
 ATGGaGTaWT GGAATTTAAT TTGTnTTAAT AAAGCACTCG TTAATAACT TGnAAAAATG 780  
 CTAACGATAA ATGTAATAGG ATAGTCAAAG CG 812  
 15

(2) INFORMATION FOR SEQ ID NO: 287:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1732 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

ATnnATTATT ATTACTGCTA TTTTAAATTT TAAAAAATGC TTTTGATTAT ATTCAACAnT 60  
 TTGTATAAAA TTAAATTTGC TTTTGATTAA AGCATGAAAA TTGTAATCAA ACCATAAATT 120  
 30 GTCGTATGAT GTAGTTAGAA TTTTAAATG CAGGAGGTCa AGTATATGAC TGAAATaACA 180  
 TTCAAAGGTG GACCAATCCA CTTAAAAGGT CAACAAATTA ATGAAGGTGA TTTTGCACCT 240  
 GATTTTACAG TGTTAGATAA TGACTTAAAT CAAGTAACAT TAGCAGATTA TGCTGGTAAA 300  
 35 AAGAAATTAA TTAgtGTGGT ACCATCAATT GATACAGGTG TTTGTGATCA GCAGACTCGC 360  
 AAATfCAACT CTGATGCTTC TAAAGAAGAG GGGATTGTGC TTACAATTTT AGCAGACTTA 420  
 40 CCATTTCGCAC AAAAAAGATG GTGCGCTTCA GCAGGTTTAG ACAATGTCAT TACATTAAGT 480  
 GACCACCGTG ACTTATCATT TGGTGAAAAC TATGGCGTTG TTATGGAAGA ACTTCGCTTA 540  
 TTAGCTCGTG CAGTATTTGT ATTAGATGCA GATAATAAAG TTGTTTATAA AGAAATCGTT 600  
 45 AGTGAAGGTA CTGATTTCCC AGATTTTGAT GCTGCTTTAG CTGCATACAA AAATATTTAA 660  
 TCATTAAAGA GATAAATCTT AAAATGTATA CATCGTGTCC ATCGTTGTCA ACAGCATTAA 720  
 AATAGAATTG TTTTCTATGA TTGCTAAGAC CTATGGGCAC TTTTATTGG AGAGGGACGA 780  
 50 ATATGGCAGA ACAACAAACA ATTATGGAAC GCTTGTTTCA TACATTAGAT GAAAAAGCTA 840  
 AAACATTAAA TAATGAAAAT GGCCaAAGTT TTATTGAAAA TCTTGGGCTA GCAATGGAAC 900

55

CATTCCAATT TGCATATTTA AGTTTAATGC aGGAAGAAAA GATAcAAGCA AATCATCAAA 1020  
 TTACACCAGA TTCAATTGGA TTGATACTAG GATTTTTAGT TGAGCGTTTT ATGAACAACC 1080  
 5 AAGAAGAATT ACATATTGTT GATATTGCAA GTGGTGCCGG TCATTTAAGT GCTACTGTAA 1140  
 AAGAAGTGTT ACCTGraAtT GcGGTTATGc ATcATTTaAT TGaAGTTGAt CCAGTTTTAT 1200  
 CACGTGTTAG TGTACATTTA GCAAACCTTCT TAGAAATTCC TTTCGATGTG TATCCTCAAG 1260  
 10 ATGCCATCAT GCCACTACCA TTAGAAGAAG CAGATATCGT TATTGGTGAT TTTCCAGTAG 1320  
 GCTATTATCC AATTGATGAA AGAAGTAAGG AGTTTAAGCT AGGTTTTGAA GAAGGACATA 1380  
 GTTATTCACA TTATTTATTA ATAGAACAAG CAATAAATGC ATTAAAAGAT GCTGGATATG 1440  
 CCTTTCTAGT GGTACCAAGT AATATTTTTA CAGGTGAACA TGTAACACAG CTTGAAAAAT 1500  
 ATATTGCAAC AGAGACAGAG ATGCAAGCAT TTTTAAATTT ACCACCAACT TTATTTAAAA 1560  
 20 ATGAAAAAGC GCGAAAATCT ATATTAATTT TACAAAAGAA AAAATCGGGT GaAACAAAGC 1620  
 CAGTTGAAGT ATTATTGGCA AATATTCCTg ATTTcAAAA TTCCTTCACC AATTTCCAAG 1680  
 GATTTATGGA CAGAGTTAAA ATCCAGTGGG ATGGGACCAC CAAATCGTCC TA 1732

25 (2) INFORMATION FOR SEQ ID NO: 288:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2779 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

AAAAGAACTA GCTAAACGCA AGCAAGAAGC TATTAGTAGA ATTAAAGACT TTTCAAATGA 60  
 AAAAATAAAT AGTATTCGAA ATAGTGAAAT TGGCACAGCT GATGAAAAAC AAGCAGCAAT 120  
 40 GAATCAAATT AACGAAATTG TGCTTGAAAC AATTAGAGAT ATTAATAATG CGCATACATT 180  
 ACAGCAAGTT GAGGCTGCAT TGAACAATGG TATTGCTCGA ATTCAGCAG TACAAATTGT 240  
 AACATCTGAT CGTGCTAAAC AATCGTCAAG TACTGGAAAT GAATCTAATA GCCATTTAAC 300  
 45 AATTGGTTAT GGAAGTCAA ATCATCCATT TAACAGTTCT ACTATTGGAC ATAAAAAGAA 360  
 ACTTGATGAA GATGATGACA TTGATCCACT TCATATGCGT CACTTTAGTA ATAATTTCTG 420  
 TAATGTTATT AAAAACGCTA TTGGTGTGGT GGGTATCTCT GGTTTACTAG CTAGTTTCTG 480  
 50 GTTCTTCATT GCCAAACGTC GTCGTAAAGA AGATGAAGAG GAAGAATTAG AAATAAGAGA 540  
 TAATAATAAA GATTCAATAA AAGAGACTTT AGACGATACA AAACATTTAC CACTTTTATT 600

55

## EP 0 786 519 A2

	AAATAATGGC GAGTCACTCG ATAAAGTTAA ACATACGCCG TTCTTCTTAC CAAAACGTCG	720
	TCGTAAAGAA GATGAAGAAG ATGTGGAAGT TACAAATGAA AACACAGATG AAAAAGTGTT	780
5	GAAAGATAAC GAACATTAC CACTCTTATT CGCAAAACGA CGCAAAGATA AAGAGGAAGA	840
	TGTTGAAACA ACAACTAGTA TTGAATCTAA AGATGAGGAC GTTCCTTTAT TATTGGCTAA	900
	AAAGAAAAAT CAAAAAGATA ACCAATCCAA AGACAAAAAG TCAGCATCAA AAAATACTTC	960
10	TAAAAAGGTA GCAGCTAAAA AGAAGAAAAA GAAAGCTAAG AAAAATAAAA AATAATTTGT	1020
	TTCTTTGATA AATAGaGGAG CACCGATTGA CATCACATCA GTCGGTGCTC CTTTTATTTA	1080
	TTCTTTTTAA TTAATTTATA CAATGCCTGT TGAGCGTGTT GATTGCTTC TTTGTTTTGT	1140
15	TCTCTCGGTA TCCATTTAAC AAATAATAAA TCAAATCTT TTTCAAATAT TTCTATTTGA	1200
	TCAAAATAAG GTTTGAAATT TCGTTTTTTC ACATAACCAG CTTCAATGCT ATCTGCAATT	1260
20	AGCTTTGAGT CTGTATATAA TAGTGCGTTT TGAACATTTA ATTCACGTGC ATGTTCTAGT	1320
	GCATAAATAC ATGCAGCCCA TTCTGCAGTG TGGTTATCCA TTTGCGCTAA CTCATGTGTA	1380
	TATGTATAAT GCTGCTCATC TTCTTTGATT ACAATGGCAC ATGTACTTAT GCCTGGATTT	1440
25	cCTTTCGTCG CAGCATCAAA ATTTATTTTC GCCATAATAA ACCTACTTTC TATTCAATAC	1500
	TTAGTTAAAG TTAATATTAC TGTAATACAA AATATGTTGG GTAATCCATT AAAAAACACG	1560
	CATCACTTAA ATAAGTAACA CGTGTTTAAA ATACTCGCTG ATTCAAAGAT GATTTTCTAA	1620
30	TACGTALACT GTaATATACT TCCTAAAAAA ATCATCTTCA GGCTGGGACA TAAATCAATG	1680
	TTCTATGCTC TACGATGTTA TATTGGCAGT AGTTGACTGA ACGAAAATGC GCTTGTAACA	1740
	AGCTTTTTTC AATTCTAGTC AGGGGCCCCA ACACAGAGAA TTTGAAAAG AAATTCTACA	1800
35	GGCAATGCGA GTTGGGGTGT GGGTCCCAAC ACAGAAGATG ACGAAAAGTC AGCTTACAAT	1860
	AATG <sup>1</sup> TGCaAG TTTGGGATGG GCCCCAACAA AGAGAAATTG GATTCCCAAT TTCTACAGAC	1920
40	AATG <sup>2</sup> CAAGTT GGGGTGGGAC GACGAAATAA ATTTGCGAA AATATTATTT CTGTCCCACT	1980
	CCCTTAAAAC TTATTCTTTT GTGTAGTAAG TCGTTAATA GCCTTGATCT AACTTATCAA	2040
	TCTTACCTTT ACGATAAAAT GATTAGCAA TATATCCAAA TGGTACATTG AAAACTGTTG	2100
45	AAGCTAATTT TAATACGTAC GTTGTAATAA ATATTTCAaA TACAa <sup>1</sup> MTGa CCAGGTAAAC	2160
	TTCCGATAAA TGCGATAGCT ACAAATAAAG CTGTATCAAT TATTG <sup>1</sup> rGCTT AAAAATGTAC	2220
	TACCATATGg CACGGATGAA AAACGTTTTa TCMGa <sup>1</sup> ACTAA ATACTTTTTT AATTAGTGAA	2280
50	AAGATAAATA CATCAATATG TTGACCAATA ATATATGCGA CGATTGAGCC TAAAGCAATG	2340
	CGTGGCACAA CATCAAAGAT TCGGTGTAAT GCTTTTTGTG CCATATCTTC TGGTGCAGGA	2400
55		



CAAACTGCTC TTTTGTCAAC TCTACGCCCA TAAATATCGT TTAATATATC TGTTGCTAAA 2520  
 TAAATAGAAG CAAACATGAC ATTACCTAAA GTTGCTGAAA TACCAAAGAT TTCTACAGTT 2580  
 5 TTAATCACTT GTATGTTGGC AATGATTGTG CCAATTGCAA CCCATGCAAT TAAACCTTGT 2640  
 TTACCAAAAA AGCGATACAT AAGTACCATA AGCACGAACG TTGCAATAAA CGTAACTAGT 2700  
 CCTAAAAATT CATTATACAT ATTAAAATGT CCTCCTAAAT TTTGATCATG CGGGTGTTTA 2760  
 10 GAAACCGCTC AATAAATAA 2779

## (2) INFORMATION FOR SEQ ID NO: 289:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1999 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

ACTGATGTGC GTTCATCAAA AACAATATAA TCAAATTCAT TTTCATCAAA TGCTTAAAA 60  
 25 TTATCATCTC TAGATAATGT TTGAATGGTT GCAAATAAAT ATTTGGCATC GACATCTCTA 120  
 TGTTTTCTCG TCAACAATCC AAAATCACTA TCATTTTTTA TAGGTAATAC TTTTTTAAAT 180  
 TCTTCCTTAG CTCTATTTAA AATCCCCTCA TTATGAACAA TAAATAAAAA TTTATTAGGG 240  
 30 TTTACTTCTC TAACATCTAA TGCACATAAA ATCGTTTTAC CTGTACCAGT TGCAGATATT 300  
 ATTAACGCCT TATCTTTGGC TTTATCCCTA ATAGCTTTTA ATGACCTTAA TGCTTCTGCT 360  
 35 TGCATTAAAT TGGGTACAAT TTCCACTGAT TTTTTCACCT TATCAGCTAG CAGCATTTGA 420  
 GTTTGTTCAA CCTCCGCTAA TTTTCTAAG GAACGGTACT CAAATGATTC TTTATATGAA 480  
 TTAATCCATT GCTCAGTCAG TGGGGTACTC TTTTGCCATA ACAAGTCAAA TTCACTTTTT 540  
 40 ACACTATCAA CTAAATCGCC ATTTTTCATA GTAGACAGTA AAACATTATG CTCATAATTA 600  
 ACCTTTAACG CATTAGATGT TAAATTAGAG CTTCTTATTA CCATAGAACT ATAATCCTTA 660  
 TGCTCAAAAA TATATCCTTT GGCATGGAAT CCAGCAATAT CAGTTAATCT TACCTCTACA 720  
 45 TTTTTTAATT TAAGTAATTC TCCATACATT TTAGGACTAT TAAACCCTAA GTAATTAGAT 780  
 GTTAATATTT TCCCTTTAAC ACCCTTATTG CTTAAATCTA ATAGTTGAGC CTTTAAGCTG 840  
 GcTAAACCGC TTTCTGTTAT AAAAGCCACA GAAAAATAAA ACGTTTCACA TTTTGAAGT 900  
 50 TCATCTATAA TTGTTGAAAG AACTTTTTCA TTTTATTAT TACTAAAAG CTTCCGGTGTA 960  
 TAATTCCTT TATGAGAAAT ATGTTTGTCT ATAAACCCTT TATGTAAAGA TTGATTGAAA 1020

55

CGCTGGGGCC CAATTTAATT TATCAAGTTC GTTTATTGAC AACCATTCAA TACTCTTATG 1140  
TTCAGTTAGA GTTGGTAACT CTTTGTTTAA AGTACATTTG TATGTTGTTA ACCTAACAAAT 1200  
5 TCCAAAATCA TATTCATGTT CTGTAGTTAT AACTTTGTCT CCAACAATTA AATCACATTT 1260  
CATTTCTTCT CTAATTTCTC TAATCAAAGC GTCTTTTTCa GTTTCATTCT TTTCAACCTT 1320  
ACnGCCAGGA AATTCACACA TTAAAGGCAG ACTCATTTTT TCACTTCTCT GTGCACAAAG 1380  
10 AATTTTGTTA TCAGAAAAAA TAATAGCTCC TACTACATTG ATTACTTTTT TCATAAGACT 1440  
CACCTTCAA TTTAAATCA TCTTAATTGT TATTCTATCA AAAATTACAA AACTATATAT 1500  
15 AAATCAATAT TAAAAATTAA TATTTTACAT TCACATGAAC GCTCTACTCC ATGCATTTTC 1560  
ATACACATCT ATTATATAAT ACTTGTGAAA AGTATTGTCT TGGGGCTGTG TTTTTTTACT 1620  
TTTGGGGCGT ATTTCTTTAT AATTCATTAC ATAAATGTAA GGGCTTTAGT TTTTCATGTTT 1680  
20 TATTAAGTCT AACTGAGATT TTGAAAGGAT GTTTAGCAAC AATGGATAAA GAATTATGGA 1740  
TAGAACGAGc TAATGATAGT TTAGTTAAAC ATTTTTATGa GCAGCAATCT GACATTGAAC 1800  
AGCGAGAAGG TTTTGAAAGT AAATTAACAT TTGGTACTGC GGGTATACGC GGAAAATTCG 1860  
25 GTCTTGGTGA AGGTGCACTT AATAAGTTTA CTATTGAAAA ATTGGCATTa GGTTTAGCGC 1920  
GTTATTTAAA TGCCCAAACA AACAGTCCAA CAATAGTCAT TCATTATGAT ATTAGACATC 1980  
TTTCcAACTG AATTCGCCC 1999

30 (2) INFORMATION FOR SEQ ID NO: 290:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1933 base pairs  
(B) TYPE: nucleic acid  
35 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

GATGACTTTT CCCCCTCATA ATCTTCATGG TCCAGGCGTC CATTAATGCG TCAAAGGATG 60  
GCACATTTTA CCTGGAACAA ATGATTCATA TGGTTCATAA AAATCACGCG TCGTAATATA 120  
45 ATCTTCTAAA TCAAATGCAT AGAAAATCAT TGGCTTTTTA AATACTGCAT ATTCATATAT 180  
TAAAGATGAA TAGTCACTAA TTAATAAATC TGTATGAAC AGTATATCAT TAACTTCTCT 240  
AAAGTCAGAA ACGTCAACAA AATATTGTTT ATGTTTGTCT GCAATATTAA GTCTATTTTT 300  
50 CACAAATGGA TGCATTTTAA ATAATACAAC CGCGTTATTT TTTTCGCAAT ATCTTGCTAA 360  
ACGTTCAAAA TCAATTTTGA AAAATGGGTA ATGTGCTGTA CCATGACCAC TACCTCTAAA 420

55

TTGTTTGATC TGTGTCGCAT AAGCTTCATC AAATAGTACA TCAGTACGTG GAACACCTGT 540  
 AGGCACTACA TTTTCTCTT TAATACCAA TGCTTCAGCG TAGAATGGAA TATCGGTTTC 600  
 5 AGATGATACA TAAGCTTTTG TATAGCTACG ATGATTTAAT GAATCAATAA ATGGTCCACC 660  
 CTTTTTACCA GTACGACTAA AGCCAACTGT TTTAAAGGCA CCAACGGCAT GCCATACTTG 720  
 AATAACTTCT TGAGAACGTC TAAAACGCAC TGTATAAATC AATGGGTGAA AGTCATCAAC 780  
 10 AAAGATGTAG TCTGCCTTCC CAAGTAAATA TGGCAATCTA AACTTGTCGA TGATGCCACG 840  
 TCTATCTGTA ATATTGCTT TAAAAACAGT GTGAATATCA TACTTTTAT CTAAATTTTG 900  
 ACGTAACATT TCGTTATAGA TGTATTCAA GTTCCAGAC ATCGTTGGTC TAGAGTCTGA 960  
 15 TGTGAACAAC ACCGTATTCC CTTTTTCAA GTGGAATAA TTCGTCGTAT TAAATATCGC 1020  
 TTTAAAAATA AATTGTCTTG TATTAAATGA TTGTTGCGG AAATACTTAC GTAATCTTT 1080  
 20 ATATTACGA ACGATATAA TACTTTTAA TCCGGAGTC GTTACAACAA CATCAAGGAC 1140  
 AAATTCATTA ACATCGCTAG AAATTTGAG TGTAACAGTA TAAACCGTTT TCTTCGAAAT 1200  
 GCCGCCTTT CTAAATCTT TTAGGTAAGT CTGCAATAAG AAATTGATTT TACCATTTTG 1260  
 25 TGTTCCTAAT TCGTTGTATT CTCTCTCTG TTCTGGCTT AGATTTTGAT ATGCATCATT 1320  
 AATCACATCT GGGTTTAACT GTGCAATATA ATCAAGTTCT TGCTCATTC CTAATAAGTA 1380  
 CTTATCTTCA GGTAAGTAAT AACCATTATC TAAGATAGCT ACATTGAAAC GACAAACGAA 1440  
 30 TTGATTCCCA TCTATTTTGA CATCATTCGC CTTCATTGTA CGTGTCTCAG TTAAATTTCT 1500  
 TAATACAAAA TTAATATCTT CTAAATCTAG GTTTTCACTA TGTCTTCAA CGAATAACTG 1560  
 AACACGTTCC CAATAGATTT TATCTATATA TATCTTACTT TTAACCAACG TTAATTCATC 1620  
 35 CTTTTCTATT TACATAATCC ATTTAATAC TGTTTTACCC CAAGATGTAG ACAGGTCTGC 1680  
 TTCAAAAGCT TCTGTAAGAT CATTAATTGT TGCAATTTCA AATTCTTGAC CTTTAAACAA 1740  
 40 CGCTAATTTA nCTACAATAT CTGGGTATTG AATGTATAAG TCTACAACAT CTTGGAAATC 1800  
 TTTTGAACCA CTTGACTAC TACCAATCAA CGTTAGTCCT TTTTCCAATA CTAGACGTGT 1860  
 ATTAACCTCT ACTGGGAACT CACTTACACC TAACAGTnCA ATGCTTCCTT CTGGTGAAAT 1920  
 45 GTAATCGATC ATT 1933

(2) INFORMATION FOR SEQ ID NO: 291:

(i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 2049 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

	nGtnCGGnCA GATATATTGG TGGTCTTTAG TAAGTGATC AAATTCATCA GATGTCAAGG	60
5	GCATGTTATC ACCTCCTTAG GTTGATAACA aCATTATACa CGaAAGGAGC ATAAaCAAaT	120
	GAACACAaGA TCAGAAGGAT TGC GTATAGG CGTCCCACAA GTTCTAGCA AAGCTGATGC	180
	TTCTTCATCC TATTTAACGG AAAAGGAACG TAACTTAGGA GCGGAAATAT TAGAACTTAT	240
10	TAAAAAAAGT GATTACAGCT ACTTAGAAAT AAACAAAGTT TTCTATGCAT TAGATAGAGA	300
	ACTTCAATAC AGGGCGAATA ATAACAACT TTAaCATTa TCTAAAGGAG TGATAGAGAT	360
15	GCCAAAAATC ATAATACCAC CAACACCAGA AAACACATAT CGAGGCGAAG AAAAATTTGT	420
	GAAAAAGTTA TACGCAACAC CTACACAAAT CCATCAATTG TTTGGAGTAT GTAGAAGTAC	480
	AGTATACAAC TGGTTGAAAT ATTACCGTGA AGATAATTTA GGTGTAGAAA ATTTATACAT	540
20	TGATTATTCA GCAACGGGAA CATTGATTAA TATTTCTAAA TTAGAAGAGT ATTTGATCAG	600
	AAAGCATAAA AAATGGTATT AGGAGGATTA TCAAATGAGC GACACATATA AAAGCTACCT	660
	ATTAGCAGTG TTGTGCTTCA CGGTCTTAGC GATTGTACTC ATGCCATTGC TGTACTTCAC	720
25	TACAGCATGG TCAATTGCAG GATTCGCAAG TATAGTGACA TTCATATTTT ATAAGGAATA	780
	CTTTTATGAA GAATAAAAAA ACTGCTACTT GTTGGAGCAA GTAACAGTGC AAGATGAGCA	840
	ATTGTCTTAA ATAATTATAT AAGGAGTTAT TAATATGACC TTACAACAAA AAATACTATC	900
30	ACATTTTGCA ACATATGACA ATTTCAATTC TGATGATGTT GTTGAACTT TTGGGATATC	960
	TAAAAACAT GCAAAATCCA CACTTTCAAA ACTTAAGAAA AAAGGAAAGA TTGCAATGGA	1020
35	AAGTTGGGGT GTCTGGCGTG TTATTGAATC GCAATTGCAT TTAAGTGTAG TCGAACGTAA	1080
	AAAAGAAATT TTAGAAGAAC AATTTGAATT GTTAGCAAGA TTAAATGAAC AAAGTGATGA	1140
	CCCTAGAGAA ATAGAAGAAC GTATCAAGTT AATGATTCGT CTAGCTAACC AATTTTAAGG	1200
40	AGGAGTTAAT CAATGGCAGT ATTAGAAGGT ATTTTGAAG AATTAAACT ATTAAATAAG	1260
	AACTTACGTG TGTTAAACAC TGAAGTATCA ACTGTAGATT CATCAATTGT ACAAGAGAAA	1320
	GTAAAGAAG CACCAATGCC AAAAGAAGAA ACAGCTCAAC TGAATCAAT TGAAGAAGTT	1380
45	AAGGAACTT CTGCTGATTT GACTAAAGAT TATGTTTTAT CAGTAGGAAA AGAGTTCCTT	1440
	AAAAAAGCAG ACACTTCTGA TAAGAAAGAA TTTAGAAATA AACTTAACGA ACTTGGTGCG	1500
	GATAAGCTAT CTAATATCAA AGAAGAGCAT TATGAAAAA TTGTTGATTT TATGAATGCG	1560
50	AGAATAAATG CATGAAGCTA GATCACTCAA ATAGAGCTCA TGCAAAGCTT AGTGCAAGTG	1620
	GAGCAAAACA ATGGCTAAAC TGTCCACCGA GTATTAAGGC AAGTGAAGGT ATTGCAGATA	1680
55		

GTCTTAAATA TGAAGGCCTA ACACAGTTTG AGTTTAATAA AGCTTTTCAA AATTATAAGC 1800  
 GAAATCAATA TTACAGTGAA GAGTTGCGCG AATATGTTGa AGAGTACGTA GCTAATGTAG 1860  
 5 AAGAAAAGTA TAACGAGGCT TTGaGTAGAG ATGACGATGT AATAGCTTTA TTTGAAACAA 1920  
 AATTGGATyT AGGTAAATAC GTCCCTGAAT CTTTGGTay TGGTGATGTC AtTATATTTT 1980  
 CAGGTGGTGT ACTTGAAATT ATTGACCTTA AATACGGTAA AGGCATTGAA GTTTCAGCTA 2040  
 10 TAGATAATC 2049

## (2) INFORMATION FOR SEQ ID NO: 292:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 942 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

ATGATGTTTC TATATTCGTA TTAGGAAAAC CTGTTGGTAT TACAACAAAC GCCCTAAAAT 60  
 25 TACAATACCG CTGCTCCTAT ACCAATTGCA ACAACAGTTT TAACTGAAAT ATCTTGTTTT 120  
 TTCATCTTCA TTA CTCTt ACATAAAAAA TTCATTATAT TGATGGTGCT TTAGATAAAT 180  
 GAATCGTCCA ATCATTCCA GTACCAATAT GATATAAATC TGAAAATGAG TCTTGATTGA 240  
 30 CTGCTACACC AATATTTACT AGCGAGTTAA CATACACAAG AGGTTACCCC ACATTAACAT 300  
 CTGCAAACGA TCGCGCAAAT TTAATAATAT TTTGATAGAC TTTCTTATCT TGATGATAAA 360  
 TTGTTACCAC CAAATTATTA CCATGAACAA TTTCCAAGGA TTTTAAGAAT GCCAATGGAA 420  
 35 TATTTGTCCA TAATGACCCA AATCTGATAT CTAAATATC AATGCTTCCT GTAACAGAAT 480  
 CCTCATT TTT TGTCATTCT CTTATTCTA ATGCCTCAAT ACTATCAACA TTAATTGCCT 540  
 40 GACCGAGACG TTCAAACGCT ATCTTATTTG CAGCTAATCT CGCACCATTG TATGCATAAA 600  
 CATCTCTACC ATGAAAAGTA TGACTTTCTT CCGAATGAGG CAATCGGCTT TTCACTTCAT 660  
 CAATTTTCGAT AACTTTTTTA ATACCTTCGT AATGTTAAT ATGACTTAAA GAACCATTAT 720  
 45 CAGGTGTAAT AATGTAATGA CCTGAATATG TTAAGCAAGC AATGTCCGCC TATCACTACC 780  
 TACACCCGGG TCTACCACTG ATACAAaC TGTGCCTTTA GGCCAGTATT TTACAGTTTG 840  
 ATATAAACGA TATGACGCTA CCCAAATGTC ATACGGTGGT ATATCATGCG TTAAGTTTTT 900  
 50 AACACGTATA TCATCATTAC AGTATATGCA ACTCCATACA TT 942

## (2) INFORMATION FOR SEQ ID NO: 293:

55

(A) LENGTH: 1268 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

10	TGTAATAAAA TTTTATGnAA CATGCTGnGA TGCTACCATG AACCTTCTTC ATTTCTTTAT	60
	GTGAGATTGT GaAATTAATC AAATAATAAT ACGGTGGATA CTTTCCTAAT TTACGATATT	120
	CCATTTCTCTG ACGATAAAAT GTTAAATAAT CATTITTTTTG AACATCCAAT ATTGAATAAT	180
15	GATCTGGATT ATACGTTTGA ATGATGACTT GACCTGCCTT TTCATGACGA CCAGCTCTAC	240
	CAGCCACTTG CGTTAATAGT TGATAAGTAC GTTCGCTCGC CCGAAAATCA GGTAAATTTA	300
	ACATTGTATC TGCATTGAGC ACACCAACTA AAGTAATATT TGGATAATCT AATCCTTTTCG	360
20	CAATCATCTG AGTACCTAGT AAAATGTCAC CGTTACCTTT TTCGAATTCA GTCAATAACT	420
	TTTCATGTGC ACCTTTCTTT GAGGTTGTAT CTACATCCAT CCTAATTATG CGCGCATCTT	480
	CAAATTCTTG TTGCAATAGT TCTTCAACTT TCTGAGTACC AGTACCTACT TGTGGAATGT	540
25	GTTCACCTCTC ACAATTTGGA CATTGATTCG GTGGCGTCTC TTGGTAACCA CAATAGTGAC	600
	ATTTTAATAA GTCTGTCGTT TTATGATACG TTAATGAAAT ATCACAGTTT GGACATTGCG	660
30	GTACATATCC ACAATCCCGA CATAACATAA ACGATGCATA ACCACGTCGA TTTAAAAATA	720
	AAACAACCTTG TTCCTGTGCA TCTAATCTTA ATTGTATGGC TTCACGTAAA TCTTTTGAAA	780
	ACATTGACCG ATTACCTTCA CTCAATTCTT CACGCATGTC TACTATATCA ATTTCAGGTA	840
35	AAGCTTGTTG GTTCACTCTG TTTGGTAATG ATAGCAAATG ATAAACGCCT TTTTCAGCTC	900
	GTGCATAACT TTCAAGACAT GGTGTTGCAC TTCCTAAAAT GACTGGACAG TGATGATATT	960
	CAC <sup>H</sup> TCGCCA TTGGGCAATT TCTCTAGCGT GATATCTCGG ATAATCTTCT TGTTTATATG	1020
40	TAGATTCATG TTCTTCATCA ATGATGATTA ACCCTAAATT TTTGAAAGGT GCGAACACAC	1080
	TTGACCTTGC ACCAACACTT ACTCTCGCAC GACCATCCCT AATTTTTTGC CATTTCATCAT	1140
	AACGTTCCCC ATTAGATAAG CCAGAATGTA ATACAGCAAC GTCATCACCA AATCGACGTT	1200
45	TGAAGCGTAA AACCATTGTC GGTGTTAGAG CGATTTCAGG AACTAACATC ATCGCCTGTT	1260
	TTCTTGG	1268

50

(2) INFORMATION FOR SEQ ID NO: 294:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 629 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

5 TACCACCAAA TAATATATTA GCTGGCATT TAATAACATT TAaATTGTC ATGATATCAT 60  
 CAATAAAATG TTGAAACTTC GTAATTTAC CTTCATAATC ATCAATTGCT GCTAATTGCG 120  
 CACTCGATGC TTGCTGATCT AAATTTAAAA TATTCGACAT GCGTTGACTA TAATAAACTA 180  
 10 AATGTTCTAT TAAGCCATCG TCACTCTTTT CCTTTGGTGC TGACATGACA GCGATACGTT 240  
 TCAAAGGATA GTGTTGCGCC AATTTTAATG TCATTAATCC ACCTAAAGAC ACACCCGTTG 300  
 CACTGATAGA TTCATAACCT TCATTGACTA AAAATTGGTA AGCTTTCTCA ACTTCTTCCC 360  
 15 ACCAATCATC TACATTATAT GTCATGAAAT CTTTCAACAA TAAACCATGA CCTGGATAAT 420  
 TCGGTGCATA ACAACTAAAT CCTTGGTCAT TTAACTCAGC TGCAAGATGC TTCACATCCC 480  
 GATTGTACC TGTAATGAA TGTAATAATA ATATCGCATG TCCATTTGTG CCTTTTAAAT 540  
 20 ACGtGGACTC GGTGTTTTAA TTCTCATTTT TctaTATAcC TCCACTAtGT CTAAAGakGT 600  
 TkGCTAAACG CGTTGtCGTC GATGATTAA 629

## 25 (2) INFORMATION FOR SEQ ID NO: 295:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2817 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

35 TATGAAAGTA ATGAATGGTA ATATTATTAA ATTTGATGGA AAAGTAGATA TTGATAATGC 60  
 AGATFAATATC GGTTTTTTAA TTGAGCATCC TAAATTATAT GATAATAAAT CAGGATTGTA 120  
 40 TAACTTGAAA TTATTTGCAC AAGTATTAGG TAAGGGTTTT GATAAAGCAT ACACAGACAA 180  
 AATTATAGAT GCATTTGGTA TGAGACCTTA TATTA AAAAG AAAGTTAAGA AATATTCAAT 240  
 GGGGATGAAG CAAAAGTTAG CAATTGCAGT ATCTTTArTG AATAAACCTA AATTTTTAAT 300  
 45 CTTGGATGAG CCTACAAATG GKATGGATCC AGATGGCTCa ATTGATGTGC TGA CTACAAT 360  
 TAAGTCTTTA GkAAATGaAC TTGATATGAG AATTCTaATA TCAAGTCATA AGTTAGAAGA 420  
 TATTGAATTA ATTTGTGATA GAGCTGTATT TTTAAGAGAC GgnCATT TTG TTCAAGATGT 480  
 50 AAACATGGAG GAAGGTGTTG CATCTGACAC AACGATAGTT ACTGTTGATC ATAAAGACTT 540  
 TGATAGAACT GAAAAATATC TTGCAGAGCA TTTCCAATTA CAAAATGTCG ACAAAGCAGA 600

55

	ATTAGATATT TATCCGAAAT ATATTGAAAC ACGTAAAAGT TCATTGCGTG ATACGTACTT	720
	CAATATAAAT CAAAGAGGTG ATAAATAATG AGAATTTTAA ATTTAGTTAA GTATGATTTT	780
5	TATAGTATAT TTAAaArTCC TTTAACATAT TTAGCGaTAC TAGTCGTATC TAGTTTGATT	840
	GCAACTCAAA GTATACTTAT GGCAAATTCG ATGGATAACC CGAAACATAT TATTGTCTAT	900
	GGATCTGTAT TTGCTGCAGC AAAATGGTTA TTGTTAATAA TTGGATTAAT GTTTGTTGTT	960
10	AAGACAATTA CGCGTGATTT TTCACAAGGT ACAaTTCAAC TATATATGAG TAAAGTTAAA	1020
	ACACGCGTTG GATaCATTAT TTCGAAAACA ATTTCAATTa TTTTAATTTc AATATTATTT	1080
15	GCATTAATTC ATTATGTGAT TTTGATTGTT GTGCAGGCAT CTAGTAATGG AAAAAATTG	1140
	GCGTTTTCTA AATATGTAGA TAATTTATGG TTCTTCCTAA TCTTTTTACT ATTCTTTGGC	1200
	TTGTTTTTAT TCTTAATCAC ACTTGCATCA CAAAAACAG CAATGATATT TTCATTAGGT	1260
20	GTATTTTTAG TACTCATTGT ACCGTTTATT AAACCTTTTA TTACATTTAT CCCAAGATAC	1320
	GGTGAAAAAG TTTTAGATGC TTTTGATTAT ATCCCTTTTG CTTACTTAAC TGATAAAATG	1380
	ATTAGCTCTA ACTTTGATTT TAGCAATTGG CAATGGGTAA TTTCATTAGG TTCTATAGTG	1440
25	ATATTCTTCA TTTTGAATAT CTTATATGTC GCTAAAAAG ACATyTAATA AAAATAATTT	1500
	TGAGGTTGGG AATTTTAAAT TTTCCCAACC TCAAAGTTTG TCTTATTGTA AATTTATTTA	1560
	TTTTCTAATT TATTTAGGAT GGAATTATAA ACTGCTTTCC AAAATGAAGC GTCAGTTTAA	1620
30	TAGCGGTTTG ATATAACTAA GTGTGTTTCT TTTTCTAAAT CTGCATAGTC TGGATGATCT	1680
	TTGCTCGGTA ATTTATCAGC ACGAACATCA GTTACAAATT TTTGGACTTC ATTTGCTCTT	1740
35	GGTCCCCAAA CTGTTTCTTG TTCGAATTGA TCATTCAAGA ATACGAAGAT AGGAATTGCA	1800
	CGTGATTTAC CATTTGTTAA ATATTGATCG ATCAGTTTTG TATCATCATC TCTATGGAAC	1860
	ACGCGTACTT CTAAATTTAA TGCTTCACTG ATGTGTTTTA GAATTGGGAG ATTCATCATT	1920
40	GCATCTCCAC ACCAGTCTTC AGTAATTACT AATACTTTAG AATAATTCAT CTCTTTTATT	1980
	TTTTTGATGC GTGAATCATC TTCTGGTAAC TCAAATGATT GATAGATACT GAGAACGGTA	2040
	TCTTGATTTG TCTTCATTCC ATCAATGTAT TCATTTAAGG GTTGGCTATT TTGAAATAA	2100
45	GTTTCTAAAT TTGTCATTGT AAAAACCTCC TTTAGCATTT ACAACATTAT ACCAATTTAT	2160
	AGTAATAAAA GGTAATGAAA TAAATTAATT GCAAATTCCT TGTTAATTTT TGTTAAGGAT	2220
	GAAACGGGAA GCACCTTATG CTATATTTAA ATAAGTACAA AGAAAGGGTG ACATCAGTGC	2280
50	GTATTCAAAA TCGCTGGGTT GTGTTTATAT TATTTTAAAT CTGTTCTTTT GGTGTATTAA	2340
	TTGGTTTATA TCAATATCGT CATACGAAAA CTGTGGATTT GTCTAATCTT GAAATAAACG	2400
55		



TTGATCGTTT TAAATTTTAT AACAGTAAAG CTCACCCTGA TCTTACCGTT AAAGTGAGAG 2520  
 AAAAGGATAA CATCGTTAAG GGGATAATAT TAGTAAGAGA TGAAAAGATA CATACTAATT 2580  
 5 TTGATGGGGG AATTGGTTCG CCGATAAATA ACGCgATTGA AAATCTTGGa TTCgGATATA 2640  
 AAAGrACaAA AGTTGGcAAT GrtTkCtCAT CgGTAAAGTA TATTGATAGA GATAACCATT 2700  
 TAAAATTAAA CTTACTTTTAT CAAGATTTAG AAATTAAACG TATTGAATTT TTTAGTAAAT 2760  
 10 AGCTTTAGGT CTTAAAGTTw TAAAAAACGA ATGAaTAATT TTATTGGGAT GAGTGAC 2817

(2) INFORMATION FOR SEQ ID NO: 296:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 1607 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

TCTGTAAAA TGATTTTTCT TTTAnAAAGG CCGnAAATCA ATGTTGATT nTTATTTGCA 60  
 25 TTATGGTCTC GATATTGGTA GAATATCAAA TGGTTAAATG AGAAAACTT GGAGGTGCTC 120  
 ACATGTCATC AATCGTAGTA GTTGGGACAC AATGGGGAGA CGAAGGAAAA GGAAAAATAA 180  
 CGGATTTCTT GGcAGAACAG TCAGATGTTA TCGCGCGTTT TTCAGGTGGT AATAATGCAG 240  
 30 GCCATACCAT TCAATTTGGC GGAGAAACAT ATAAATTACA TTTAGTACCA TCTGGTATCT 300  
 TTTACAAAGA CAAATTAGCG GTAATCGGTA ACGGAGTCGT TGTTGATCCA GTTGCACTAT 360  
 TGAAAGAATT AGACGGATTA AATGAACGTG GCATTCCCTAC AAGTAATTTA CGTATATCTA 420  
 35 ATCGTGCGCA AGTGATTTTA CCATATCACT TAGCACAAGA TGAATATGAA GAACGTTTAC 480  
 gTGGgGACAA TAAGATTGGT ACAACTAAAA AAGGTATCGG TCCAGCATAT GTAGACAAAG 540  
 TTCAACGTAT CGGTATTCGT ATGGCAGATT TACTTGAAAA AGAAACATTC GAAAGATTAT 600  
 40 TAAATCAAA CATTGAATAT AAACAAGCAT ATTTCAAAGG TATGTTTAAC GAAACATGTC 660  
 CATCATTTGA TGATATCTTT GAAGAATATT ATGCAGCAGG TCAACGTCTA AAAGAATTTG 720  
 TAACAGACAC ATCAAAAAATC TTAGACGATG CATTTGTAGC AGATGAAAAG GTACTTTTCG 780  
 AAGGTGCGCA AGGTGTAATG TTAGATATCG ACCATGGTAC ATATCCATTC GTTACATCAA 840  
 GTAATCCAAT TGCAGGTAAC GTTACTGTTG GTACAGGTGT AGGTCCTACA TTCGTTTCAA 900  
 50 AGGTAATTGG TGTATGTAAA GCTTATACAT CACGTGTTGG TGATGGTCCA TTCCCTACTG 960  
 AATTATTCGA TGAAGATGGA CATCATATTA GAGAGGTTGG TCGTGAATAC GGTACAACAA 1020

55

TAAGTGGTAT TACAGATTTA TCTATTAAC TCAATCGATGT TTAAACAGGC CTAGACACAG 1140  
 TGAAAATCTG TACAGCTTAT GAATTAGACG GTAAAGAAAT TACTGAGTAC CCAGCAAAC 1200  
 5 TAGATCAATT AAAACGTTGT AAACCAATCT TTGAAGAGTT ACCAGGTTGG ACAGAAGACG 1260  
 TAACAAATGT GCGTACTTTA GAAGAATTAC CTGAAAATGC ACGTAAATAT TTAGAGCGTA 1320  
 TTTCAGAATT ATGTAATGTA CAAATTTCTA TCTTCTCAGT TGGTCCAGAT AGAGAACAAA 1380  
 10 CAAACCTATT AAAAGAATTG TGGTAGAACT TTATATAAGT CATAACAAT GATTATAAAT 1440  
 ACATGAGCCT TCTATCTTTA TTGGTAGGAG GCTTTTGTTA TGCTTGCTTC TGTATCGATT 1500  
 CGATTATTTA GATAAAAAAT ACTAACGTAA AGGCGATATT TGCTAGTCAT AATTTAGAAG 1560  
 15 rTTAgatGat AtTtAACGAA AAtTAaGATG anATActTGA ATGGTAA 1607

(2) INFORMATION FOR SEQ ID NO: 297:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3055 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

TTAGAAGTAA GCACTTTAAT ATCTTTACCT ACCCATGTAC CAACACCTTC TTTAGGCTTT 60  
 30 GGATTTTCAG CATGGTTATT TGATTATTTC ACCTGTTTCA ATCTATTAC TTTATTACCT 120  
 TTATTGGGGT TCTCTAATAC ATCAAATTTT AATCTCGGTG AATAAAAAAG ATATATTAAA 180  
 AATCCAATA AAAATAATAC ACCGACGACT CTTATAATTA ACTTTTTCAT CAATCAACCA 240  
 35 CCTAAAAAGT ATTAATACTA TTGTAAAAAA CAACACATTA ATTAGCAAAT TTTCAACACT 300  
 GACATAACTG TGTGTTTTCG ATAAAAACAA ATCATCTTCA GGCATAATTT TAACGTCATA 360  
 40 AAGATTTTCC CTACACTCTA TATCATAACC TATCTTTATG TTTTCAGGTT GAATTCATT 420  
 ATTAAGATTA AAATATGTAT AAAAAAATGG ACAGTTAAGG TATCAAATGA ATACCATCAA 480  
 CTGTCCGACT ATTCTTCATC AAAAAACCTG ATAAAAACAA TTGCCTTATC AGATTAGTAT 540  
 45 CATTTGTATA AGCATATTAA TGACCAAATG TTGCTTTAAT CAGTGATGTC GTTTCTCCAC 600  
 CTGGATATAA TACATATAAT AATAAGTAAA CTGCTACACC TGTAATTGCA GTACAGAACC 660  
 ATATAACTGA AGCGAATGGC CCGAATTTGC GGTGTACATT ATATTTATCT TTAAATGCAG 720  
 50 TAATAATTG AACTAGGCCT AGAATACCAC CAATTGTTGC TAAATTAATG TGGAAAAATA 780  
 AGAAAATCGT ATAATATTTT TTAATTGATG CTGGTCCGCC AAATGCTGTA TTACCGATAA 840

55

	TAACATTTTT GTGTTTATTT ATTTCCCTTT TCCAAATAAG TCTCCAACCA ATGGCAATTA	960
	AAATTGCACT AATGACAATA CATGTCGTAC TAATCGTTGG TAAAATTGGA ACGCCCATAT	1020
5	TTTTCATCCT AACTTAATTA ATCTAGATCA AAGTAAGTAA TGAAACAATC ACAGCTAACA	1080
	CGAAAAAGAT CACTAAATAA TTTAGTGAAT ATATAAACAT TTGTGTTGCC CATTTTGT	1140
	GATCTGAATT TTTCTTAAAT GTTGTTAAAC CTAATGCAAT CCATCCTAAA TTTAATAAGG	1200
10	TrGCTAACAC TaCGAATACG ACACCTAAAT TTATTAGTAA TAAAGGTACT GGCAATAAAA	1260
	TAATCAACCA GATAACATA CTGACACGTG TACGTTTAAA GCCCTTAACT GATGGTAACA	1320
15	TTGGAATATT TGCAAGTGCA TATTCATCTT TACGTTTAAAT AGCTAAGGcA TAAAAATGAA	1380
	TTGGtTGCCA ACmAAATACA ACTAAAAACA GCGCAATCGC TGTTAAACTA ATTTGTCCTT	1440
	CAATTGCAAC CCATCCAATT AGTGGTGGTA CTGCTCCAGG AAAACTCCCA ATCACTGTGT	1500
20	TCCATGTTGT ATGTCTTTTA GACCATATTG AGTAATAAGA CACATAACCT ACAATCCCCA	1560
	TAAGACCAAG TACGCCTGAT GGTATATTCA ATAAAAACAA ACAAATTTCT CCAACTAACA	1620
	TCATACCAAA ACTTAATAGT AATAAATTTT GATCTGTAAT TCTATTATTT ACAGTTGGTC	1680
25	TATTTTGT	1740
	ATGCGCCACC CATAATTAAA GTAGATCCAA ATAGCATTA TAAAATTGA GGTATTGATG	1800
	ATAAGAAGGA ATGATTGTG ATTACAACCTG CTAGCCATGC GCcCGCAAAA GCCGGAATTA	1860
30	AGTTACCTTG AACAAGTCCC ATTTTAATTA TCTGTTGCAA TTCTTTGAAG TTAACCTG	1920
	TAATATTTTG TGACmAGTA TGCTCTTGC TCATAATCCC CCTCCTTAAA TTTGTTTATA	1980
	TAAGATTATG ATATCTTAGA TTGCATAAAA AGACTAGGTT TAATAAAATT AAATTGTGAC	2040
35	AAATTAAACGA CAAGAGAAAA TGTCATTTT GTGACACAAA TAACATTTAA TTTATTGCTA	2100
	TAATGTATAT GTTAGAAAAT TTTAATAAGT AGAATCATGC ATCTAAAAGA GATTAATATT	2160
40	TAAGCTTCAA ATTTGAGTAA ACGTGGATTA CATAATTATC CCAATAAAAA AATCATTACG	2220
	ATTAAGTTCT TTTTATGTCG TCCACATACA ATACTTGTA AATTAAATCA TATTCCTGC	2280
	GTTGGATCCC ATCTTTTCAT ATCCTACAAT CAGGTCTATT TATAGTATCA TCTCAAATCC	2340
45	GGCTATTAAAT TCTAATTCTC AGTGATGCGT TTTTTTATGA TGGGGTGAT AAATTGTTTG	2400
	GCAAAAAGAA TTTAAATGG TTAGGTGTCG TAGCAACGTT AATGATGACA TTTGTACAAC	2460
	TTGGTGGAGC CTTAGTTACC AAAACCGGAT CAGCTGATGG TTGTGGTTCT TCTTGCCAC	2520
50	TATGTCATGG TCGTTGATT CCAGAAITCT TTCCTATTGA TACGATTATT GAGTTAAGTC	2580
	ATAGAGCCGT TTCAGCTTTG TCTTTATTAA TGGTCTTATG GTTAGTTATC ACTGCATGGA	2640

55

TATTGCAAGC ATTAATCGGA GCTGCTGCTG TTATTTGGCA AAAAAACGAT TACGTTTTAG 2760  
 cATTGCACTT TGGTATATCA TTAATCAGTT TCTCATCTGT ATTTTAAATA ACATTGATTA 2820  
 5 TTTTCTCTAT AGATCAAAAA TATGAAGCTG ACGAATTATA TATCAAAAAG CCATTAAGAC 2880  
 GTTAAACATG GTTAATGGCA ATCATCATTT ATTGTGGTGT TTATACTGGT GCACTAGTGa 2940  
 GACATGCGGA TGCAAGTTTA GCATATGGTG GTTGGCCATT GCCATTACCA CGATCTTGTA 3000  
 10 CCACATTcAG GAACAAGATT GGGTTCAACT CACGCATCGT ATCAnGGTCn nTTAA 3055

(2) INFORMATION FOR SEQ ID NO: 298:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 748 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

TTCTTATTTA AAGAAGTCAT TTTTAGAAAT TGTGAGACT TTAAAAAATG ATCCGTATAA 60  
 25 AATAACACAA TCTTTTGAAA AATTAGAGCC TAAATATTTA GAGCGATATT CAAGAAGAAT 120  
 TAACCATCAG CACAGGGTCG TCTATACCGT AGATGATCGA AATAAAGAAG TATTAATACT 180  
 ATCGGCATGG TCACATTATG ATTAATGAAT ATTCAATATC TGAATAACTT TAATGATAAG 240  
 30 TTAATTAAAG AAAC TAGTAT TTAAGGTAG GGAAATAGC GACGTTAATG CGTTGTTATC 300  
 TCTACACTTT TTAATTTTAT AATAGCGCAA GACTAAACAG ATTGAAATTA GTAACAATAA 360  
 AAGAATAACG TATTATAATA AGGAATTTTA AATTGTGACT TTTTCGGAAT ATTAAATTTT 420  
 35 AGAAATATGA GGTTTTTAAG CGGATTCCTC ACAAATTTT AAAAATATTT AAGCCTGAAA 480  
 ATGATAAAGC GGTAGGGAAC GTTTTTCTGA AaGTTAGTGA TACAATAGTT TTAAGTTGAA 540  
 40 ATACAGGAGG ATGAATAACA TGAATCAGTC AGTCAAATTA CTTAAACATT TAACAGATGT 600  
 AAACGGCATT GCTGGTTATG AAATGCAAGT TAAAGAAGCA ATGCGTaaCT ATATAGAGCC 660  
 TGTCAGTGAT CaaATTATTG AAGATAACTT GGGTGGCATT TTTGGAAAGA AAAATGCTGA 720  
 45 GAATGGTCAA TACTCAATTA TGGAnTTC 748

(2) INFORMATION FOR SEQ ID NO: 299:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4718 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

	ACCTCCGAAT AATTGATTCC ATTAACTTTT TTTGTG <sub>a</sub> AAA AtTTAAACAG GCGAAGTCTT	60
5	CAATAAGTGA AGATGTCCAA ATTATAAAAA ATACATTCCA AAAAGAAAAG TTAGGTACAG	120
	TAATTActAC ycTGGCGCAA GTGGTGGTGT TACGTATAAA CCAATGATGA GTAAAGAAGA	180
	GGCGACTGAA GTTGTTAATG AGGTCATTAC TCTATTAGAA GAGAAAGAAC GTTTGTTA <sub>c</sub> C	240
10	TGGCGGATAT TTATTTTTAT CAGATTTGGT AGGTAATCCA TCGCTACTAA ACAAAGTTGG	300
	TAAGTTAATT GCCAGTATTT ACATGGAAGA AAAATTAGAT GCTGTTGTTA CCATTGCGAC	360
15	AAAAGGTATT TCATTGGCAA ATGCGGTTGC TAATATTTTA AATTTACCAG TAGTAGTGAT	420
	TAGAAAAGAC AACAAGGTGA CTGAAGGTTT TACAGTTTCA ATTAATTACG TTTCAGGATC	480
	TTCAAGAAAA ATAGAAACAA TGGTACTTTC GAAGAGAACT TTAGCAGAAA ATTCAAATGT	540
20	TTTAGTTGTC GATGATTTTA TGAGGGCTGG TGGCTCTATT AATGGTGTTA TGAATTTAAT	600
	GAATGAGTTT AAAGCCCATG TAAAAGGGGT ATCAGTACTT GTAGAATCAA AAGAAGTTAA	660
	ACAAAGATTG ATTGAAGATT ATACTTCCTT AGTGAAATTA TCTGATGTAG ATGAATATAA	720
25	TCaAGAGTTT AACGTAGAAC CTGGCAACAG TTTATCTAAG TTTTCATAAA AGGAGTTTTA	780
	GTATTATGAA AATCATTAAAC ACAACAAGAT TACCGGAAGC ACTTGGACCA TATTCGCATG	840
	CAACAGTTGT GAATGGTATG GTTTATACTT CTGGTCAGAT TCCATTGAAT ATTGATGGAC	900
30	ATATCGTAAG CGCTGATGTT CAAGCACAGA CAAAACAAGT TTTAGAAAAT TTAAGGTTG	960
	TTTTGGAAGA AGCAGGATCT GATTTGAATT CTGTTGCGAA AGCGACCATT TTCATTAAAG	1020
	ATATGAATGA TTTCCAAAAA ATAAATGAAG TGTATGGTCA ATATTTTAAT GAACACAAGC	1080
35	CAGCGCGTAG TTGTGTAGAG GTTGCGCGTT TGCCAAAAGA TGTGAAAGTA GAAATTGAAT	1140
	TAGTAAGTAA AATTAAGGAA TTATAATTTT CGATTAATAT GTTTAATCAA GCTTCTAAAT	1200
40	AAAACAGAGA GATATATACT ATAGGGGGGC TCACTACATG AAAGTGACAG ATGTAAGACT	1260
	TAGAAAAATA CAAACAGATG GACGAATGAA AGCACTCGTT TCCATTACAT TAGATGAAGC	1320
	TTTCGTAATT CATGATTTAC GTGTAATTGA AGGAAACTCT GGCTTGTTCTG TTGCAATGCC	1380
45	AAGTAAACGT ACACCAGATG GTGAATTCCG CGACATCG <sub>c</sub> g CATCCTATTA ATTCAGATAT	1440
	GAGACAAGAA ATTCAGATG CAGTGATGAA AGTATATGAT GAAACAGATG AAGTAGTACC	1500
	AGATAAAAAC GCTACATCAG AAGATTCAGA AGAAGCTTAA TCAATTTTAT ATTTAGCGAT	1560
50	GTAATACATT TGCAATAAGT TGATTTGATA CTGTCGATAA AGCATAAAGC TTTGTCGGCA	1620
	GTTTTTTTTAG TTTGTATTAA TGTTTTTTTA TTTTAAATGA AAGGCTAATA AATATATACG	1680

55

	TGaTGCTCGT	ATTTTGAAG	TAAGAAAAA	GTTGTTTTTA	AAATTACAAC	GAATTAAAAA	1800
	CAATGCCTTT	TATATGTTGA	AAGAGTATTG	CaGATTAAAT	TaTAATAATG	ACGAaGgTAA	1860
5	AATTTAATGG	GGGTTAATGT	TCATGCGAAG	ACACGCGATA	ATTTTGGCAG	CAGGTAAAGG	1920
	CACAAGAATG	AAATCTAAAA	AGTATAAAGT	GCTACACGAG	GTTGCTGGGA	AACCTATGGT	1980
	CGAACATGTA	TTGGAAGTG	TGAAAGGCTC	TGGTGTCGAT	CAAGTTGTAA	CCATCGTAGG	2040
10	ACATGGTGCT	GAAAGTGTA	AAGGACATTT	AGGCGAGCGT	TCTTTATACA	GTTTTCAAGA	2100
	GGAACAACTC	GGTACTGCGC	ATGCatGCAA	ATGGCGAAAT	CACACTTAGA	AGACAAGGAA	2160
15	GGTACGACAA	TCGTTGTATG	TGGTGACACA	CCGCTCATCA	CAAAGGAAAC	ATTAGTAACA	2220
	TTGATTGCGC	ATCACGAGGA	TGCTAATGCT	CAAGCAACTG	TATTATCTGC	ATCGATTCAA	2280
	CAACCATATG	GATACGGAAG	AATCGTTCTGA	AATGCGTCAG	GTCGTTTAGA	ACGCATAGTT	2340
20	GAAGAGAAAG	ATGCAACGCA	AGCTGAAAAG	GATATTAATG	AAATTAGTTC	AGGTATTTTT	2400
	GCGTTTAATA	ATAAACGTT	GTTTGAAAAA	TTAACACAAG	TGAAAAATGA	TAATGCGCAA	2460
	GGTGAATATT	ACCTCCCTGa	TGTATTGTCG	TTAATTTTAA	ATGATGGCGG	CATCGTAGAA	2520
25	GTCTATCGTA	CCAATGATGT	TGAAGAAATC	ATGGGTGTAA	ATGATCGTGT	AATGCTTAGT	2580
	CAGGCTGAGA	AGGCGATGCA	ACGTCGTACG	AATCATTATC	ACATGCTAAA	TGGTGTGACA	2640
	ATCATCGATC	CTGACAGCAC	TTATATTGGT	CCAGACGTTA	CAATTGGTAG	TGATACAGTC	2700
30	ATTGAACCAG	GCGTACGAAT	TAATGGTCGT	ACAGAAATTG	GCGAAGATGT	TGTTATTGGT	2760
	CAGTACTCTG	AAATTAACAA	TAGTACGATT	GAAAATGGTG	CATGTATTCA	ACAGTCTGTT	2820
	GTTAATGATG	CTAGCGTAGG	AGCGAATACT	AAGGTCGGAC	CGTTTGCGCA	ATTGAGACCA	2880
35	GGCGCGCAAT	TAGGTGCAGA	TGTTAAGGTT	GGAAATTTTG	TAGAAATTAA	AAAAGCAGAT	2940
	CTTAAAGATG	GTGCCAAGGT	TTCACATTTA	AGTTATATTG	GCGATGCTGT	AATTGGCGAA	3000
40	CGTACTAATA	TTGGTTGCGG	AACGATTACA	GTAACTATG	ATGGTGAAAA	TAAATTTAAA	3060
	AcTATCGTCG	GCAAAGATTG	ATTGTAGGT	TGCAATGTTA	ATTTAGTAGC	ACCTGTAACA	3120
	ATTGGTGATG	ATGTATTGGT	GGCAGCTGGT	TCCACAATCA	CAGATGACGT	ACCAAATGAC	3180
45	AGTTTAGCTG	TGGCAAGAGC	AAGACAAACA	ACAAAAGAAG	GATATAGGAA	ATAATCATTT	3240
	ACGTATTTAA	AATGGCTAGG	ATAAAAGGAT	AATCCTATGT	AATATTAATG	TAATCTTTAT	3300
	GATTTAATGA	TTCGCATAGT	AATGGAGTTA	CATyTTATAT	ATAATAGTAA	TTGCGTAAGT	3360
50	AAATAATTGG	AGGACTATAA	ATGTTAAATA	ATGAATATAA	GAATTCGTCA	TTAAAGATTT	3420
	TTTCATTGAA	AGGAAACGAA	GCATTAGCGC	AAGAAGTTGC	TGACCAAGTA	GGAATTGAAC	3480

55

GTATTCGTGG TTGTGACGTA TTTATTATTC AACCAACATC ATATCCTGTG AATCTACATT 3600  
 TAATGGAATT ATTAATTATG ATTGATGCTT GTAAACGTGC TTCTGCAGCA ACAATCAATA 3660  
 5 TTGTAGTGCC ATATTATGGA TATGCAAGAC AAGATAGAAA AGCCCGTAGC CGTGAGCCAA 3720  
 TCACTGCTAA ATTAGTTGCA AACTTAATCG AAACAGctGG CGCAACTCGT ATGATTGCGT 3780  
 TAGACTTACA TGCACCACAA ATTCAAGGAT TCTTTGATAT TCCAATTGAC CACTTAATGG 3840  
 10 GTGTGCCAAT TCTTGCTAAA CATTTCAAAG ATGATCCGAA TATTAACCCA GAAGAATGTG 3900  
 TCGTTGTTcA CCAGACCATG GCGGsnTTAC ACGTGCACGT AAATTAGCTG ACATTTTAAA 3960  
 15 AACTCCAATT GCAATTATAG ATAAACGTCG TCCTAGACCA AATGTTGCTG AAGTGATGAA 4020  
 CATIGTTGGT GAGATTGAAG GACGTACGGC AATTATTATT GACGATATTA TTGATACAGC 4080  
 AGGTACAATC ACTTTAGCTG CACAAGCATT AAAAGATAAA GGTGCTAAAG AAGTATATGC 4140  
 20 TTGTTGTACA CACCCTGTTT TATCAGGACC GGCTAAAGAA CGTATCGAAA ATTCTGCTAT 4200  
 AAAAGAATTA ATCGTAACAA ACTCAATTCA TTTAGATGAA GATCGCAAAC CATCTAACAC 4260  
 TAAAGAATTA TCTGTTGCTG GTTTAATCGC ACAAGCTATC ATTCGTGTAT ACGAAAGAGA 4320  
 25 ATCAGTTAGC GTATTATTTG ACTAATATTT AAAAGGCGTT TGACGAACAT ATTCCAAACG 4380  
 TGTATAATAG TTTCGTTCTG GATTATACGA ATAAATAAAC ACTTGCAAGC AACGATGATG 4440  
 TTGATGGGTA AGTGAGGTGC TCGTTTTGAG CAAAATGAA AGGTGGAAAT GAGAATGGCT 4500  
 30 TCATTAAAGT CAATCATCCG TCAAGGTAAA CAAACACGTT CAGATCTTAA ACAATTAAGA 4560  
 AAATCTGGTA AAGTACCAGC AGTAGTATAC GGTTACGGTA CTAAAAACGT GTCAGTTAAA 4620  
 GTTGATGAAG TAGAATTCAT CAAAGTTATC CGTGAAGTAG GTCGTAACGG TGTTCATCGAA 4680  
 35 TTAGGCGTTG GTTCTAAAAC TATCAAAGTT ATGGTTGC 4718

(2) INFORMATION FOR SEQ ID NO: 300:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3181 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

AAAATGCATA TAAATACATA TTAAGGAGGA TTTTATGAAT TTTCTTAAAC CTGCAAAGCA 60  
 50 TATTAAGCCT TTGCCAGAAA ATCAGATAGA TGATACCTAT AAACGATTAC GTCTCCAAGT 120  
 ATTTCTTGGT ATTTTCATCG GTTACGCTGG GTACTATTTA TTACGTAAAA ACTTTTCGTT 180

55

	TGCTGTTTCC ATCGCATATG GATTTAGTAA GTTCTTTATG GGTACTGTAA gTGATCGGAG	300
	CAATGCTCGG ATATTCTTAG TTCTTGGATT AGCACTCACT GCTATCGTCA ATTTGTAAAT	360
5	GGGATTTGTA CCGTTCTTTA CATCAGGTAT CGGTATTATG TTTGTCCTAT TATTCTTAAA	420
	TGGATGGTTT CAAGGTATGG GCTGGCCACC TTCAGGCCGT GTTCTCGTTC ACTGGTTTAG	480
	TGTAAGTGAA CGCGGAAGTA AGACTGCCCT TTGGAACGTT GCGCATAATG TTGGTGGAGG	540
10	TATTATGGCA CCTATTGCTG CTTGGGGTAT TACAACAACCT GCATTTATCA ACTTTGGTTA	600
	TTTAAAAGGT TTCGAAGGTG TATTCAITTA CCCTGCACTC TTAGCACTTA TCATTGCCGC	660
	AATTCATAC GTATTGATTA GAGACACACC TCAATCTCAA GGTTCACCTC CAATCGAAAT	720
15	TTATAAAAT GACTTTGCTA CAAGCGATAA GAAAACATTA GAAACAGAAT TAACTACAAA	780
	AGAAATTTTA TTAAATATG TACTGAACAA TAAATGGGTA TGGGCAATTG CCTTTGCAAA	840
	TATATTTGTT TATTTCTGTC GTTATGGTGT ACTTGATTGG GCGCCAGTCT ACTTAAGTGA	900
20	AGAAAAACAT TTCGACTTAA AAGCATCAGG TTGGGCATAC TTCTTATACG AATGGGCTGG	960
	AATTCCTGGT ACATTATTAT GTGGTTACAT TTCTGATAAA TTATTCAAAG GTCGTCGTGG	1020
25	ACCTGCAGGT TTCTTCTTTA TGTTAGGTGT CACAGTATTT GTATTAATTT ATTGGTTAAA	1080
	TCCTCCAGGC AATGCTTGGT TAGACAATGT CTCATTAATT GCCATTGGTT TCTTAATATA	1140
	TGGACCAGTT ATGTTAATG GTTTACAAGC ATTAGATTAT GTACCTAAAA AAGCAGCTGG	1200
30	CACAgcAGCT GGATTAAACAG GATTATTTGG TTATCTGTTT GGTGCTGTAA TGGCCAACAT	1260
	CGTCTTAGGT GCTGTAGTTG ATAAATTCGG ATGGGATGTC GGTTTTATTT TATTAACAGC	1320
	AATTAGTGTG TTTGCAATGT TGAGCTTTAT CCTCACTTGG AATAAAGTAG GACAAGAAAC	1380
35	CGTTCATCAT TAAATGATAA AAAATAAAGT CATATGGTTA TCTTATCGAA AGATGATATA	1440
	TTCTCTCTT ATAAGTTCAA CCATATGACT TTTTATTAGT ATTCAAAAAA ATATTTACAT	1500
	TGCCACTTTT GTGTTTGCCC TGCTGTTTTA TTCAATTGAT TACACCACTT AGGATAAACT	1560
40	CTAAAAGCCA TTTTCCCTTG ATACTTAGAT GAAGCTAAGA TACCTTTTTC TACCAATAAT	1620
	TCCCTAGGAA ATAGAAAGTA GCCATTTAAT TCATCATCGA TAACAGCAAC AATCAGGTAA	1680
45	TCAGCAAACT CTTCAATTTT GTATGGCCGA TTATAATTAT CtTCGTCTTT TGTCCAACAT	1740
	GTCAAGAAAT ATCCCGATTG AGTCGGTGTT TTCTTAGCTA ACCTACTTTG ATATGTTTCC	1800
	TCTTTAAAGC TAAATGTTAA TGCyTCGTAA TCTTGATTAT ATTTTCTTC AGTTAAATCT	1860
50	TTAACTTCTG ATTCTTCACT AAAAATATTT TTCAACAGTA TTTTAGATTT ACACATATGC	1920
	CAATCTCACT TTATTTTTTC ATAATCGTAT CATATATTTA TTTTTTTCGA AAAATACACT	1980

55



TGTATCAACT GTCAAACGAT CACGTAATAA ATAGACGATT AACATTGcTG CTAAAGCACA 2100  
 AAGTGACTCA GCAATTAATA ATGACCAAAT GACACCTGTT AGTCCAAACA AAGCATTTCAT 2160  
 5 AATAAATAAT ACTGGAATGA TAATTGCACC TTGTAAAATG GCCATAATTG TAGCACCACG 2220  
 ACCTTGCCCA GTCGCTTGAA GCATACCAGT AAACAAGAAA CCTATACCAT TTAATAATAA 2280  
 TGATGCCATT GTTACTTTCA AAATAAATGc CGCCATCTCA ACAATGGCTT GATCAGTAGT 2340  
 10 AAATAGTCCG ACCATATGAT GTCCAATTGT AAATACAGCA CTCATACATA CAACAAAGAT 2400  
 AACGCCGATA GACATGATAA CTGCTTTGAT AACGTCTTTC ATACGGCCTT TATTTGCCAT 2460  
 AAAGTTATAT GCAATTAGTG GTACAACACC TTCACATAAT CCCATGATAA TAAGTTCTGG 2520  
 15 aAATTGCACA AGTCTAAATG AGATACCATA ACTTGCaATC GCGAAGTTTC CATAATGTGC 2580  
 TAAAAATAAA TTTAAAACTA ATCCTGTGAA TCCCATTAG ATACTCATT AAAATGCAGG 2640  
 AATACCGATT TTAAAGATTT CAGAAAGCAT TTCTTTATTA GGTTCGCAA GTTTAATATT 2700  
 AACTGACACA ACGTCACTAT TTTTCATAAA ATAAATGATA AAGAACAGAG CAGCAGCAAC 2760  
 ATTACTGATT GCAGTACCCA AAGCTGcACC AACACGTTT aATCAAAAC CAAAAATTAA 2820  
 25 AATTGGATCT AAAATAATAT TTAAGCCTAC ACTAGCTAAC ATACCAATCA TAGAAACCAT 2880  
 TGGTGCCCCA ATTGCACGTG CAAATTGTTC TAATATGAAG AACAAAATTA CAAAAGGTGC 2940  
 ACTTAAAAAC ATTACTTTCA AATAATTACT TGTTAAAGCT AACGTTTCAC CTCTCGCCCC 3000  
 30 TAAATTGCT GCGATTTGAT CACTGAATGG TAAAGTAACT AAAATCACGA TAAGTCCTAG 3060  
 TGCAATACCA CCATAAATAG AGAACTACT TACAAATTTA CTCTTAmTAT AGTCTTTCGC 3120  
 ACCTAATAAA CGTGAAATAT AAGTTCCTGC ACCAACGCCA AATAAATTAC CTAACCCCAT 3180  
 35 T 3181

## (2) INFORMATION FOR SEQ ID NO: 301:

(i) SEQUENCE CHARACTERISTICS:  
 40 (A) LENGTH: 4029 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

TCATCACCTT CAAATAAATA GGCATGCGAT AATTTATTTG AATGATATGC ATTCGTCAAT 60  
 50 TGTGCTGTT CATCCATTCA TTACAACTC CCTTTTGCTT TTATATAAAA AGGTTGCCAA 120  
 AGAGCAACTA ATTACTGCAA CAATAGCGAC CATGATTTGT GATTATTTTT AGAACCACAA 180

55

	TTATTAATAC ACTATTTTAA GTATCTCTAA TTGATGATTG TATTATAGAA TTAAAATTGA	300
	TGGAATGCAT CAACTGGCAT AACAAATACA GTAGCACC GC CAACTTCAAC TTCAACTGGA	360
5	TATGGAATGT ACGAATCCGC ACTGCCTCCC ATAGGTGTAA TAGGTGAAAC CAACTGTTCT	420
	CTATTACCAC ACGTTTGATT AATCACAGAC AATATTTTCAT CTACACGGTC ATCATTGACA	480
	CCACATAAGA ATGTTGTATT ACCCGCTCTT AAAAACCAC CTGTTGTTGC CAATTTTGTT	540
10	GCTCTAAAGT TATTTTAAAC AAGTTGATCT GCAAGTTCCT GACTATCTTG ATCTTGACG	600
	ATCGCTATAA TCATTTTCAT TTTATAACAC CTCTTCTAAC AATTATATCA TATCTTTTCT	660
	AAATATTTGA TGATAGTTTG ATACGTGTCT TCAACAACAT TTTCAAGAGG TTGATCTGCA	720
15	TTAACGCTTT TGAACCGTTG TGATTCATTA TGAATGATTT CTTGGTAACC TTCAATTACT	780
	TTTTCGTGAA CTTTAAATCT TCTTGATCTA ATCTATTTTG ATCTCTTGAA TTTTAAATAA	840
20	TACGTTTCGC ACCTACTTCA GCACTAACAT TTAAATAAAT CGTCAAGTCT GGAATAATC	900
	CATTTATTGC AAATTCGTTT AATGCTCTTA CTCTTCAAC GCCAATCCCT CTAGCATAAC	960
	CTTGATAAGC TAATGAACTA TCGATATAGC GATCACACAA CACAACCTTA CCTTCTTTTA	1020
25	AAGCTGGTAT GACCTTTAAT ACAAGATGTT CTCTTCTAGA TGCAGCAAAT AACATTGCTT	1080
	CaGTtCTAAT GTCCATATCA TTGCCcTTCT AATACAATTT TACGTATTTT TTCACCAGTA	1140
	GGAAACCAC CTGGTTCTCT AGTCATAATG ACATCATAAT CTTTTACTAA TCTATGGTAA	1200
30	ACTTCATTAA TTACAGTTGT TTTTCCAGAG CCTTCTGGGC CCTCAAAGT TATAAAAGCT	1260
	GACATTTTAT TCATCCTCAA CTAAATTTT ATTATTTTAA ATTCTTCAA CTATCATTCC	1320
	AGTTTCCAGA TATTCATTTA CCAATCTAT CATATTTTCA GTTATTGTTT CTCCTTTGAA	1380
35	AATAATAGGA ATcCCTGGcG GATACGGGAC AATATGTCGC GCCAGAACTT TACCTTTgCC	1440
	TTTTcAAATC ACACCAAGTA ACATATTCAA AGCGTTTTGG TTTATAGTTA CCTTCAGTyG	1500
40	TTAAAAGTTG TGTTTGTTTA ACTTTAGAAA CTGATTTTTT CGGTAAAATC ATATCTTCAA	1560
	TTTTACGTAA AAGCGAATCA AATAAATACG TATCATCATG ATGCCATAAC GGCAATATCG	1620
	CTAATGCTTG ATAGTCGTCC GCTAATTCTA AATAGATATG TGCATTCTA AACCAATTTT	1680
45	GAATATCATG ACCTGTAAAA CCTTCATATT TTATCAGCAA CTTCAACGGA TCATCAACTT	1740
	GAAGCATTTT AAAACCCTTC TTCTCCAAAC ATTGATTAA TTGCGCTCTC TTATCAAAAA	1800
	ACACGGTACT ATCATATGTT TTATAGAACT CGGCAGCTGA CTCTAAACTA GCCATAATCA	1860
50	AATACGAAGG ACTAGATGTT TGGAAGTAGC TTAGATATTC TATAATAGTT TCTCTATAAG	1920
	GTGCATTTTT ATGAATATAA AGTACCGAGC CCATCGTTAA AGCTGGTAAC GTTTTATGAA	1980

55

	CAAAGTGCGC	GCCGTGTGCT	TCGTCAATGA	GTACAGGAAT	ATTTAATTGG	TGCAAAGATT	2100
	TGATAACCTC	TTCTACATTA	AATGTTTCAC	CGTAATAGTT	AGGATAAGTC	AACACAGCAA	2160
5	GTTTGTGACC	GTCATTATTC	AAACGGCTTA	AATTAACTTT	ATTATAATGA	TTCGTTAACG	2220
	GACTTTGATG	CGTTTCAATA	AAATGCCCTT	CTTGTTGGCT	AATATCGAGC	GCAATGAACA	2280
	CAGATTTATG	TACATTTCTT	GCCATTAAGA	TATCGCCTTT	TTTCTGTGAA	AAAGACTGGA	2340
10	TGACAGATAA	TATTCCTGAA	GTGGTGCCAT	TCACTAAGAA	ATAAGCATCA	TAATCTGAAT	2400
	GTTTCTCCAC	CTGCTTCATA	CTTTCCAAAA	TGACTTCTTC	AGGATGATGT	AAATCAnCTn	2460
15	AATCCAGGTA	TTTCAGTTTT	ATCCATTGTC	ATTGATAATT	GAGATAAATG	ACCGATAGTC	2520
	ATATTTTTAT	GACCCGGAAC	ATGCAAAGAA	ATCGCTTCTT	CTTGATTTAA	ACTTTCTAAT	2580
	TTATTTAAAA	TAGGTTGCTT	CATGATATAC	GCTTCCTTTA	TTTACACTGT	TTTGGAATTA	2640
20	GTTACTTTCA	AAAGTATTAA	TTATATAGTA	ACACTTCTTT	GACAAAAGTT	AGTGTTACTT	2700
	ATGCAATAGC	TTGTCTATTG	TATAATAATT	AATTTCTTTT	TTGTACTTCG	ATTTAAAAGA	2760
	TATTAGACAT	AAAATCTAAA	AACAGCAGTA	AGATGATTTA	TGATTAAAAA	CTATCTTACT	2820
25	GCTGTTCACT	TTTTATAATA	CTTCTGAATG	TCTTCACTTA	TACTTCTAGT	CACAGATTTA	2880
	AATAATCAAA	AGTGCACATT	ATTAAAAATAT	CAATTTCACA	CTCAATGCGG	CTCATCGCAT	2940
	TCATTTCTTG	TCTAGCAACG	TTCTACTCTA	GCGGAACGTA	AGTTAGCTAC	CATCCTCGCT	3000
30	AAGAACCTTT	CTTGACTTGT	GACAATCGCT	TGCTTCTTTT	CTCTCCTTCG	GCTCTCGCTT	3060
	ACTCATTTAG	CTCTACTAAA	CTCGTTGCGC	TCTTTTCTCG	TTTCGTCAGA	TTCAAACGTT	3120
	TTCACTTCGC	CAAGCCATTT	TTCTTTGTGT	TTACTTTTTA	TTTTGACGTT	TTAGACATAA	3180
35	AAAAAGAGAC	cTCACGGTCT	CAACTTGCCT	GGCAACGTTT	TACTCTAGCG	GAACGTAAGT	3240
	TGGCTACCAT	CGTCGCTAAA	GACCTTTCTT	GACTTGTGAC	AATCGCTTGC	TTCTTTCTCT	3300
40	TCCTTCGGCT	CTCGCTTACT	CATTAGCTC	TACTAAACTC	GTTGCGCTCT	TTTCTCGTTT	3360
	CGTCAGATTG	AAACGTTTTT	ACTTCGCCAA	GCCATTTTTT	TTTGTGTTTA	CTTTTTATTT	3420
	TGACGTTTTA	GrCATAAAAA	AAAGAGACcT	TGCGGTCTCA	ATGCGGCTCA	TCGCATCCAT	3480
45	TTTTTGCTTG	GCAACGTTCT	ACTCTAGCGG	AACGTAAGTT	GGCTACCATC	GwCGCTAAgA	3540
	aCCTTTCTTG	ACTTGTGACA	ATCGCTTGCT	TCTTTCCTCT	yCTTCGGCTC	TCGCTTACTC	3600
	ATTTAGCTCT	ACTAAACTCG	TTGCGCTCTT	TTCTCGTTTC	GTCAGATTCA	AACGTTTTCA	3660
50	CTTCGCCAAG	CCATTTTTCT	TTGTGTTTGC	TTTTTATTTT	GACGTTTTAG	ACATAAAAAA	3720
	AAGAGACCTT	GCGGTCTCAA	TGCGGCTCAT	CGCATCCATT	TTTTGCCTGG	CAACGTTCTA	3780

55

CATGGGAACA GGTGTGACCT CTTGCTATA GTCACCAGAC ATATGAATGT AATTTATACA 3900  
 TTCAAAACTA GATAGTAAGT gAAAAGTGGa TTTTGCTTCG CAAAACATTT ATTTTGGATT 3960  
 5 AAGTCTTCGA TCGATTAGTG ATTCGTGCAG CTCCACATGT GCACCATGCT TGCCACCTCG 4020  
 GAACCTATT 4029

(2) INFORMATION FOR SEQ ID NO: 302:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7159 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

20 GATGCAATAG TTGAGTAGTT ATAAGTAGCT ACATACAACC GCTCAAATAT AGGTTCAAGA 60  
 ACATTTTTTAA TGAAGAAAGC GATTTTCCCG CAGAGTGCAA ACGCTTGACT CAATAACGAA 120  
 ATGTCACCTA TGTATGGCGT GACTTATTAT ACAGGAGGTG CAAAGTATGT TTGCTATTAT 180  
 25 TGAAACAGGT GGAAACAAA TCAAAGTAGA AGAAGGTCAA GAAATCTTCG TTGAAAAATT 240  
 AGACGTAAAC GAAGGAGATA CTTTACATT TGATAAAGTA TTATTTGTAG GTGGAGATTC 300  
 AGTTAAAGTT GGAGCGCCAA CAGTTGAAGG TGCAACAGTT ACTGCTACTG TTAATAAACA 360  
 30 AGGTCGCGGT AAAAAAATCA CTGTATTAC ATACAAACGT CGTAAAAATT CAAAACGTAA 420  
 AAAAGGCCAT CGTCAACCAT AACTAAATT AACATCGAT AAAATCAACG CGTAATTATT 480  
 ATGATTACTG TTGATATTAC AGTTAATGAT GAAGGCAAAG TAACAGACGT TATTATGGAT 540  
 35 GGCCATGCTG aCCATGGTGa ATATGGTCAT GATATCGTTT GTGCTGGAGC TTCaGCTGTA 600  
 TTGFTTGGTA GTGTTAATGC GATTATAGGA TTGACATCTG AGAGACCAGA TATCAATTAT 660  
 GACGACAATG GTGGTCATTT TCATATAAGA AGCGTTGATA CAAACAACGA TGAAGCGCAA 720  
 40 CTAATCTTC AAACAATGCT TGTGTCTTTA CAACTATTG AAGAAGAATA TAATGAGAAT 780  
 ATTAGATTAA ATTATAAGTG AGGTGCATTC CGATGTTAAA ATTAACTTA CAATCTTCG 840  
 45 CATCTAAAAA AGGGGTAAGT TCTACAAAAA ACGGACGTGA CTCTGAATCA AAACGCTTAG 900  
 GTGCTAAACG TGCTGACGGT CAATTCGTAA CAGGTGGTTC AATTTTATAT CGCCAACGTG 960  
 GACTAAAAAT TTACCCTGGT GAAAATGTAG GTCGTGGTGG CGATGATACA TTATTCGCTA 1020  
 50 AAATCGACGG CGTTGTTAAA TTCGAACGTA AAGGTCGCGA CAAAAACAA GTTTCTGTAT 1080  
 ATGCAGTAGC TGAATAATTT TGTCTAGTTA ACACCAGAAG TGAATCTTCT GGTGTTtTTT 1140

55

	AGACGTTATA CTAAATGTGC AcTGTATAAG AGCCCCTAAT CACTAAACTA TAAGGGGGAC	1260
	AAAGGAATAC AGTTGCAGCG TTTAAAGAAT AACTGTACC ACAATTGGTG CTGAGAAATA	1320
5	TAAGTATTTT AAAGCAAAGA TTTATAAAAG TAACTGCATA AGAGCCCCTA ATTATTTACA	1380
	ATATATAAGG GGCTCTAATA TGCTATAATT ATTGGGAAAA TGAAAATTAT ATGTAAAAGA	1440
10	GGTGAGATAT ATGTTTGTCTG ATCAAGTCAA AATATCTCTT AAAGCCGGTG ATGGTGGTAA	1500
	TGGTATTACC GCATACAGAA GAGAAAAATA TGTACCATT TGGTGACCAG CTGGCGGTGA	1560
	CGGTGGTAAA GGTGCTTCAG TCGTATTGTA AGTGGATGAA GGTTTAAGAA CGTTATTAGA	1620
15	TTTTAGATAT CAACGTCATT TTAAAGCAAG CAAAGGTGAA AATGGCCAAA GTAGTAATAT	1680
	GCATGGTAAA AATGCGGAAG ATTTAGTATT AAAAGTTCCA CCTGGTACAA TTATTAAAAA	1740
	TGTTGAAACA GACGAAGTGT TAGCAGATCT TGTGAAGAT GGTCAAAGAG CTGTAGTAGC	1800
20	GAAGGGCGGT CGAGGTGGCC GAGGTAATTC ACGTTTTGCA ACACCTAGAA ACCCTGCACC	1860
	TGACTTCAGT GAAAAAGGTG AACCAGGTGA GGAATTAGAT GTATCTTTAG AATTGAAATT	1920
	ATTAGCTGAT GTAGGATTAG TAGGTTTCCC TAGTGTGGGT AAATCGACTT TATTATCTAT	1980
25	CGTTTCAAAA GCTAAGCCTA AAATTGGGGC ATATCATTTT ACAACGATTA AACCAAATCT	2040
	AGGTGTTGTT TCAACGCCTG ATCAACGTAG TTTTGTATG GCAGATTTAC CAGGTTTAAT	2100
	TGAAGGTGCA TCTGATGGCG TTGGATTAGG ACATCAATTT TTAAGACATG TAGAGAGAAC	2160
30	AAAAGTTATT GTTCACATGA TTGATATGAG CGGTTCTGAA GGTAGAGAAC CTATTGAAGA	2220
	TTATAAAGTC ATTAATCAAG AATTAGCTGC GTACGAGCAA CGTTTAGAAG ATAGACCTCA	2280
35	AATCGTAGTA GCTAACAAGA TGGATTTACC TGAATCACAA GATAATTTAA ACTTGTTTAA	2340
	AGAAGAAATT GGCGAAGATG TGCCAGTTAT TCCAGTTTCA ACAATAACGC GTGATAATAT	2400
	TGATCAATTA TTATATGCAA TAGCAGATAA ATTAGAAGAA TATAAAGATG TTGACTTCAC	2460
40	AGTTGAAGAA GAGGAGTCAG TTGGCATTAA CCGAGTATTA TATAAACATA CACCGTCACA	2520
	AGATAAATTT ACAATTTCAA GAGATGATGA TGGTGCTTAT GTGGTAAGTG GTAATGCTAT	2580
	TGAAGAATG TTTAAAATGA CTGACTTTAA CAGTGATCCA GCAGTACGTC GATTGCTCG	2640
45	TCAAATGCGT TCGATGGGTA TTGATGATGC GCTTAGAGAA CGTGGTTGTA AAAATGGTGA	2700
	TATCGTTAGA ATTCTTGCGG GAGAATTTGA ATTCGTTGAA TAGGAGCGAA ACATGATGGA	2760
	CAATAAAGAT TATAAAAAGT TTTATTTAAT TAGAGAAGAT GTCTTGCCTG AATCCGTGGT	2820
50	TAAAACATTG AAGATTAAAG ATGCCTTAAA AAGTGATCCG ACATTGTCCA TTTATGATGC	2880
	CGTTAAACAG TTTGATCTAT CTAGAAGTGC TTTTATAAAA TATAGAGAAA CGATATTTCC	2940

55

	TGTTGGTATG TTGGCACGTG TACTAGATGT TATATCAAAG TTAGAACTAT CTGTATTAAC	3060
	GATTTCATCAA AGTATTCCAA TGGAAGAAAA AGCAACAATA ACATTATCAC TGAATGCTAA	3120
5	ATCTAAAGAA ACTTCAGTAG AAGATGTTAT TGGCGCTTTG AGAAATTAG ATTATGTATC	3180
	AAAAGTAGAA TTAATTAGTA TGAGTATGTA AGGATGTGCC TATAATGTAC GCGTATGTCA	3240
	AAGTAAGTT AACACATTTA TATCCTACAC ACGTAGTTGT TGAACTGCT GGTGTTGGTT	3300
10	ATGAAATTCA AACACCAAAT TCTTATCGTT TTCAAAGCA TCTAGATCAT GAAGTTTAA	3360
	TTCATACATC TTTAATTGTT CGTGAAGATG CACAATTATT GTATGGATTT AGTAGTGAAG	3420
15	AAGAGAAAGA TATGTTCTTG AGTTTAATTA AAGTTACTGG TATTGGTCCG AAATCAGCTT	3480
	TAGCTATTTT AGCGACAAGT ACGCCTAATG AAGTAAAACG TGCCATTGAA AATGAAAATG	3540
	ATACGTATTT AACTAAATTC CCAGGAATTG GTAAGAAAAC GGCAAGACAG ATTGTCTTAG	3600
20	ATTTAAAAGG TAAAGTGAAA ATTACTGAAG AAGATAGCGA TTCATTATTA CAAGTAGACG	3660
	CTACTTCGAC GGTGCAAGAT CAATTCGTGC AAGAAGCAAT GTTAGCGTTA GAAGCATTAG	3720
	GTTATTCTAA ACGAGAGCTT GCAAAGTTG AGAAAACGTT AAATAAAAAT AAATATGACT	3780
25	CAGTTGATGA AGCTGTTAAG GCAGGTCTTC AATTAGTTGT ATCTTAATTT TAAATAGATT	3840
	AATAGGGGAA GTGTTGTCAT GAATGAGCGT ATGGTTGATC AATCAATGCA TAGTGAAGAA	3900
	ACTGATTTTG AATTGTCGCT TAGACCTACG AGATTACGAC AATATATTGG TCAAAATTCA	3960
30	ATAAAAAGTA ATTTAGAAGT ATTTATTAAA GCGGCTAAAC TTCGTCATGA ACCATTAGAT	4020
	CATGTATTGC TTTTGGCCC CCCTGGATTA GGTAAGACAA CATTATCTAA TATCATTGCC	4080
	AATGAAATGG AAGTTAATAT ACGTACAGTA TCAGGGCCTT CATTAGAAAG ACCTGGTGAT	4140
35	TTGGCTGCAA TTTTATCAGG ACTTCAACCT GGAGATGTTT TGTTTATTGA TGAAATACAC	4200
	AGACTGAGTA GTGTTGTTGA AGAAGTGTTA TACCCTGCAA TGGAAGATT CTTTTTAGAT	4260
40	ATTATCATTG GTAAAGGCGA TGAGGCTAGA AGTATCCGTA TCGACTTACC TCCATTCACT	4320
	TTGGTAGGTG CAACAACGCG AGCTGGCAGC TTAACAGGTC CACTAAGGGA TCGATTTGGT	4380
	GTGCACTTAA GATTAGAATA TTATAACGAA TCAGATTTAA AAGAAATCAT TATTAGAACA	4440
45	GCTGAGGTTT TAGGCACAGG TATTGATGAA GAAAGTGCCA TTGAACTTGC TAAACGTTCT	4500
	AGAGGGACTC CAAGAGTAGC AAATCGACTA TTGAAGCGGG TAAGAGACTT CCAGCAAGTG	4560
	AATGAAGATG AACAAATATA CATTGAAACA ACGAAGCACG CATTAGGTTT ACTTCAAGTT	4620
50	GATCAACACG GACTAGATTA CATTGATCAT AAAATGATGA ACTGTATTAT TAAGCAGTAT	4680
	AATGGCGGAC CTGTTGGTTT AGATACGATT GCCGTAACAA TTGGTGAAGA ACGTATTACA	4740

55

	GGCAGAAAAG CAACACCATT AGCTTATGAA CATTTTGCAA AGTCGAATGA GGAGAGAGAA	4860
	TAACGTGAAT ATTGAAGAAT TTGACTATGA CTTACCAGAA TCATTAATTG CTCAAACGCC	4920
5	TTTAAAAGAT CGTGATCATA GTCGTTTATT AGTCATGGAT AGAGAAACTG GTGAAATGAA	4980
	ACATTTACAT TTCAAAGATA TCATTGAGTA TTTTAGACCT GGTGATACAT TAGTGCTTAA	5040
10	CGATACGCGA GTAATGCCAG CTAGACTTTT TGGTTTAAAA GAAGAAACTG GTGCAAAAGT	5100
	TGAAATGTTA ATGTTAACTC AAATTGAAGG TAATGATTGG GAAGTCTTAC TGAAACCAGC	5160
	TAAGCGTATT AAAGTTGGTA ATAAATTGAA TTTTGGTAAT GGCAAAATTA TAGCTGAATG	5220
15	CATAAAAGAA ATGGATCAAG GTGGACGCAT CATGCGTTTA CATTATGAAG GTATTTTACA	5280
	AGAAAGATTA GATGAATTAG GGGAAATGCC ACTGCCACCA TACATCAAAG AACGTTTAGA	5340
	TGATCCAGAT CGTTATCAAA CAGTTTACGC TAAAGAAAGT GGTTCAGCGG CAGCACCAAC	5400
20	AGCAGGATTA CATTTTACTG ATGAGTTATT AATTGAAATT AAAAATAAAG GTGTTAATAT	5460
	CGCATTTGTT ACATTACATG TTGGGTTAGG TACGTTTAGA CCGGTGAGCG TAGACGATGT	5520
	GAATGACCAC GAAATGCATA GTGAATATTA TCAAATGACm CAAGAAACAG CTGATTTATT	5580
25	AAATGATACT AAGcCAAAGG ACATCGCATT ATATCAGTTG GTACAACTTC AACACGTACA	5640
	CTTGAAACAA TTCGACGCGA TCATGATAAA TTTGTGAAA CGAGTGGCTG GACTAATATA	5700
	TTTATTTATC CAGGATTTGA TTTTAAAGCA ATTGATGGCC AGATTACTAA TTTTCATTTA	5760
30	CCAAAATCAA CATTAGTTAT GCTAGTATCA GCGTTTAGTA GTCGTGAAAA TGTTCTGAAT	5820
	GCTTATAAAA CGGCAGTAAA TTTAGAATAT AGATTCTTTA GTTTTGCGCA TGCAATGTTA	5880
35	ATTATATAAA AAGAATGTGA GGATTTTGAA TATGCCTGCA GTAACATACG AACACATTAA	5940
	AACTTGTAAG CAATCAGGTG CGCGTTTAGG TATCGTGCAC ACACCACACG GTTCATTTGA	6000
	AACACTCTATG TTTATGCCAG TTGGTACTAA AGCAACCGTT AAAACAATGA GTCCAGAAGA	6060
40	GTTAAGACAA ATTGAAGCAA AAATCATTTT GGGCAACACA TATCATTTGT GGTTACAACC	6120
	CGGAAATGAT ATTATCAAAC ACGCTGGGGG ATTACATAAA TTCATGAATT GGGATGGTCC	6180
	GATTCTTACA GATTGAGGCG GTTTCCAAGT GTTTAGTTTA AGTAATTTAC GTAAAATTAC	6240
45	AGAAGAAGGC GTGGAATTTA GACATCATAC TAATGGGTCT AAATTATTTT TGAGTCCTGA	6300
	GAAATCAATG CAAATTCAAA ATGATTTAGG ATCTGATATT ATGATGGCAT TTGATGAATG	6360
	TCCACCGATG CCGTCTGAAT ATGATTATGT AAAAAATCT ATTGAACGTA CAACACGTTG	6420
50	GGCGAAAAGA TGTCTAGATG CACACCAAAG ACCTGAAGAT CAAGCATTGT TCGGCATTAT	6480
	ACAAGGTGGC GAATATGAAG ATTTAAGAGA ACAAAGTGCA AAGGATTTAG TAGAATTAGA	6540

55

AATGGTTGAA CATAcAGAGC AGTTTATGCC TAAAGATAAA CCAAGATATT TAATGGGTGT 6660  
 AGGATCTCCa GATGCGTTAA TCGAATGTAG TATTCGCGGC ATGGATATGT TTGATTGTGT 6720  
 5 CTTACCGACA CGTATTGCCA GAAATGGTAC TTGTATGACA TCGCAAGGTC GTTTAGTTAT 6780  
 TAAAAATGCA AAATTTGCAG ATGATTTAAG ACCGTTAGAT GAGAATTGTG ACTGTTATAC 6840  
 ATGTCAAAAC TATTCAAGAG CGTATATACG TCATTTAATC AAGGCAGAGG AAACTTTGG 6900  
 10 TATTCGTCTT ACTACTATTC ATAATTTACA TTTTCTGCTA AAATTAATGG AAGATATAAG 6960  
 ACAAGCCATT CGAGAAGATC GTCTTTTAGA TTCAAAGAA GAATTCTTCG AGCAATATGG 7020  
 15 ATTAAATGTT GAGAACCCAA AAAACTTTTA AGCAAGAGGA GCGTATAAAA TGCAATTTTC 7080  
 ATTACTAATA TATATAGTCG TAATTTTTCG GGTATGTAT TTCTTGATGA TCAGACCACA 7140  
 ACAAAAAC TGCGAAACA 7159

20 (2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3159 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

TCCATTTATA GAAATTTCTA AAGCAGAnAA TAAGATAGAA GATATCGGCC AAGGTGCAGA 60  
 AATCATCAAA AGAACACAAG ACATTACTAG CAAACGATTA GCTATAACTC AAAACATTCA 120  
 35 ATTTGATTTT GTAAAAGATA AAAAATATAA CAAAGATGCC CTAGTTGTTA AGATGCAAGG 180  
 CTTCAATTAGC TCTAGAAcAA CATATTCAGA CTTAAAAAAA TATCCATATA TTAAAAGAAT 240  
 GATATGGCCA TTTCAATATA ATATCAGTTT GAaAACGAAA GACTCTAATG TTGATTTAAT 300  
 40 TAATTATCTT CCTAAAAATA rAATTGATTc AGCAGATGTT AGTCAGAAAT TAGGCTATAA 360  
 TATCGGCGGA AACTTCCAAT CAGCGCCATC AATCGGAGGC AGTGGCTCAT TCAACTACTC 420  
 TAAAACAATT AGTTATAATC AAAAAAACTA TGTTACTGAA GTAGAAAGTC AGAACTCTAA 480  
 45 AGGTGTTAAA TGGGGAGTGA AAGCAAATTC ATTCGTTACA CCGAATGGTC AAGTATCTGC 540  
 ATATGATCAA TACTTATTTG CACAAGACCC AACTGGTCCA GCAGCACGAG ACTATTTTCGT 600  
 50 CCCAGATAAT CAACTACCTC CTTTAATTCA AAGTGGCTTT AATCCATCAT TTATTACAAC 660  
 ATTGTCACAC GAAAGAGGTA AAGGTGATAA AAGCGAGTTT GAAATcACTT ACGGCAGAAA 720  
 CATGGATGCT ACATATGCTT ACGTGACAAG ACATCGTTTA GCCGTTGATA GAAAACATGA 780

55



	AAAAATTAAA AGCATCACAC CTAAGTAAAC AGTTCaATCA TCTTAAAAAA TCCTGGGACA	900
	CTTCATACTT GTCTCAGGAT TTTTmACAA ATTGAATCAG CCTCATAACA TTAAATTATT	960
5	TTATCGTACA TTAAATTTAA TAATAACAAC TGATTTTTAT AAGAATAAAG TATCGAmCCA	1020
	TAGTAGATAC ACAAATAATA CAAATGAAAC AATTTAACTT GAAAGCTTAa ATAAATATTA	1080
	TCAAGTTAAT AAACAATTAA TTTTtagATG GATTcatCAA AAATCGTAAA AAAGCACAAT	1140
10	TTGTATTTTA CAAACATTAA TTAAAAAGA AAGCAAGACA TTCGTGCAAT CGGTTACCTT	1200
	AAATGTTTA CAACTGTCAA CAATACCAAG GTTTTATTAA CTATATTTCT CACAAAATTA	1260
15	GCTTTTAGCA TTCCAAACAA AAAAGGTAA ATCGAACGGA ATTATGGCAT TTTTAACTTA	1320
	ATTGTAAAAA AAGTTGATAA TGGTCAATTG TTAATGAACA GTTAATTATA ATAACGCCCA	1380
	AAATATATTA TTATTTAATT AAGTTAAATA AAATTATAGA AAGAAAGTGA AACTTATGCT	1440
20	TAAAAATAAA ATATTAACTA CAACTTTATC TGTGAGCTTA CTTGCCCCTC TTGCCAATCC	1500
	GTTATTAGAA AATGcLAAAG CTGCTAACGA TACTGAAGAC ATCGGTAAAG GAAGCGATAT	1560
	AGAAATTATC AAAAGGACAG AAGATAAAAC AAGTAATAAA TGGGGCGTGA CTCAAAATAT	1620
25	TCAATTTGAT TTTGTAAAGG ATAAAAATA TAACAAAGAT GCTTTGATAT TAAAGATGCA	1680
	AGGATTCATT AGCTCTAGAA CAACATATTA CAACTATAAA AAACTAATC ATGTTAAAGC	1740
	TATGCGATGG CCATTCCAAT aTaATATTGG TTTAAAAACA AATGATAAAT ATGTTTCTTT	1800
30	AATTAATTAT TTACCTAAAA ATAAAATTGA ATCTACAAAC GTGAGTCAGA CATTAGGATA	1860
	CAATATCGGT GGTAATTTCC AATCAGCCCC ATCACTCGGT GGTAATGGAT CATTTAACTA	1920
35	TTCTAAATCG ATTAGCTATA CACAACAAAA TTATGTAAGT GAAGTAGAAC AACAAACTC	1980
	AAAAAGTGTT TTATGGGGCG TCAAGCGAA TTCATTGCGC ACTGAATCAG GTCAAAAATC	2040
	AGCCTTTGAT AGCGATTAT TGTAGGCTA CAAACCTCAT AGTAAAGATC CTAGAGATTA	2100
40	TTTCGTTCCA GACAGTGAGT TACCACCTCT TGTACAAAGT GGATTTAACC CTTCAATTAT	2160
	CGCCACAGTA TCTCATGAAA AAGGTTCAAG CGATACAAGC GAATTTGAAA TTACTTACGG	2220
	AAGAAACATG GATGTCACTC ATGCCATTAA AAGATCAACG CATTATGGCA ACAGTTATTT	2280
45	AGACGGACAT AGAGTCCATA ATGCATTGT AAATAGAAAC TATACTGTGA AATACGAGGT	2340
	CAATTGGAAG ACTCATGAAA TCAAGGTGAA AGGACAGAAT TGATATGAAA ATGAATAAAT	2400
50	TAGTCAAATC ATCCGTTGCT ACATCTATGG CATTATTATT ACTTTCTGGT ACTGCTAATG	2460
	CTGAAGGTAA AATAACACCA GTCAGCGTAA AAAAAGTCGA TGACAAAGTT ACTTTATACA	2520
	AAACAACAGC CACAGCAGAT TCTGATAAAT TTAAATTTT ACAGATTTTA ACATTTAATT	2580
55		

ACTCAGGCTT TGTGnAACCT AATCCTAATG ACTATGACTT TTCAAAATTA TATTGGGGAG 2700  
 CTAAATACAA TGTATCTATA AGCTCACAAT CTAATGATTC AGTAAACGTC GTTGATTATG 2760  
 5 CACCAAAAAA TCAAAATGAA GAGTTTCAAG TTCAAAATAC TTTAGGCTAT ACATTTGGTG 2820  
 GTGACATTAG TATCTCTAAT GGTTTATCTG GTGGACTTAA TGGAAATACA GCTTTTTCTG 2880  
 AAACAATTAA TTATAACAA GAAAGTTACA GAACAACATT AAGTCGCAAC ACAAATTATA 2940  
 10 AAAATGTTGG CTGGGGAGTT GAAGCACATA AAATTATGAA TAATGGTTGG GGACCTTATG 3000  
 GAAGAGATAG CTTCCACCCA ACATATGGTA ATGAACTCTT CTTAGCTGGC AGACAAAGCA 3060  
 15 GTGCATACGC TGGCCAAAAC TTCATAGCGC AACACCAAAT GCCATTATTA TCTAGAAGTA 3120  
 ACTTCAATCC AGAATTTTTTA AGCGTACTAT CACACAGAC 3159

## (2) INFORMATION FOR SEQ ID NO: 304:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3821 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

30 GCAAAAACCTT TTCTCCAACCT ATGTCAGACT ATAATGGCAT GAACAAAATT GATATGATGA 60  
 ATCAAATTAA AGTTGATACG ATGTTGCATG GTTATCACGC TGGATTTTTTA TTTGCATTAT 120  
 TGATTACAGT TGTTAGTTTC TTTTGTTCAT TTATGTTACA AGGTAAGAAA AAAGAAGTTG 180  
 35 ATTCTCGTCA ATAAATATAA GTTGCTAGCT ATATAAAGCT TTTTAGCAAA AGTTCAACAT 240  
 ATTGACTTAT CCGGCATTTC AGATTAAAAT ATTTTATTC CGATTAGAAT AATAAGAATA 300  
 AGGAGATATA TTCTATGAAA AGACTTTTAT TTGTGATGAT AGCTTTCGTT TTCATATTGG 360  
 40 CTGCATGCGG AAACAATTTC TCGAAAGACA AGGAAGCTAG TAAAGATAGC AAGACAATTA 420  
 ATGTTGGGAC TGAGGGGACT TATGCACCAT TTAGTTTCCA CGATAAAGAT GGTAAATTAA 480  
 CTGTTACGA TATTGATGTT ATTAAAGCAG TGGCTAAAGA AGAAGGTTTA AAACTTAAAT 540  
 45 TTAATGAAAC TTCTTGGGAT TCTATGTTTG CAGGTTTAGA CGCAGGGCGT TTTGATGTAA 600  
 TCGCGAACCA AGTAGGTATT AATCCTGATA GAGAAAAGAA ATATAAATTT TCTAAGCCTT 660  
 50 ACACATTCTC AAGTGCTGTT TTAGTTATTC GTGAAAATGA AAAAGATATT AAAGATTTTG 720  
 ATGATGTTAA AGGTAAGAAG TTAGCACAAA CATTACATC TAATTATGGT AAATTAGCTA 780  
 AGGATAAAGG TGCTGATATT ACAAAGTTG ATGGCTTTAA CCAATCAATG GATTTATTAT 840

55

	AAAAACCTAA TGcTAAGATC AAAGCAATCA AAGGTAATGC TGAACAAAGT AGATCTGCAT	960
	TTGCATTTTC TAAAAAGCA GATGATGAAA CAGTTCAAAA ATTCAATGAT GGCTTGAAAA	1020
5	AAATCGAGGA AAACGGTGAA TTAGCTAAAA TAGGTAAGAA ATGGTTTGGT CAAGATGTTT	1080
	CTAAATCTAA ATAGCGAACA ACAACATGCG CTAGATGCTG CAAAACAAGC TTTCGGACCT	1140
	ATGCTAGAAG GTTTGGTCAA ATATTCAATT CCTATTACAT TAGTTACATT TGTTTTAGGA	1200
10	TTGATTATTG CATTATTTAC AGCATTAAATG CGAATTTCAA CGAGTAAAAT TTTAAGAAGT	1260
	ATTTACGCTG TCTATGTATC TATTATTCTGA GGAACACCAA TGATAGTACA ACTATTTATC	1320
15	ATATTTTATG GTATTCCAGA ATTAGGTAGA TTATTAACAA ATGACGCTGA CAACCAATGG	1380
	ACATTGGCAC CTGTAGTGGC TGCTATTATT GGTTTATCAT TAAATGTAGG TCGGTATGCT	1440
	TCGGAAATTA TTCGTGGCGG TATTATTTCT ATACCGAAAG GACAAACAGA AGCTGCaTAT	1500
20	TcCaTCGTA TGACGTATGG TCAAACGATA CAACGTATCA TTTTACCGCA GGCAATTCTGA	1560
	GTGTCGATTC CTGCACTAGG TAATACATTT TTAAGTTTAA TCAAAGATAC ATCATTATTA	1620
	GGATTTATTT TAGTGGCTGA AATGTTTAGA AAAGCTCAAG AAGTTGCGTC TACAACATAT	1680
25	GAATATTTAA CAATTTATGT GTTAGTTGCG CTAATGTACT GGGTGGTATG CTTTATTATT	1740
	TCAATTATCC AAGGTATCTA TGAATCTTAT ATTGAAAGAG GGTATCGCTC ATGATTCAAT	1800
	TGAACAATAT CCATAAATCA TTTAATGATG TTGAAGTCAT CAAAGGTATT GATTTATCTG	1860
30	TTGAACAAGG TGAGGTTGTA ACCTTAATCG GTCGATCTGG TTCAGGTAAA ACAACATTGT	1920
	TACGTATGAT TAATGCATTA GAAATTCCAA CTGAAGGTAC AGTTTATGTT AACGGCAAAA	1980
35	CATATACATC TAAAGATAAA AAATCACAAA TAGAAGTTCG TAAACAGTCT GGTATGGTAT	2040
	TTCAAAGTTA TAACCTTTTT CCGCATAAGA CGGCATTAGA AAATGTAATG GAAGGTCTTA	2100
	TCAÇAGTTAA AAAGTTGAAA AAGGATGAGG CACGTGGGAA ATCACTTGAG TTAÇTTGAGA	2160
40	AAGTTGGTTT AACACATGTC AAAGATCAAC GTCCACATGC ATTATCAGGT GGTCAACAAC	2220
	AACGTGTTGC TATTGTCAAG AGCACTAGCA ATGAACCCTA AAGTGATGTT GTTTGATGAA	2280
	CCAACATCTG CACTTGATCC TGAAC TTGTG AATGATGTTT TAAAGGTTAT TAAAGATTG	2340
45	GCTAATGAAG GCATGACAAT GGTCATTGTG ACACATGAAA TCGGTTTTGC TAAAGAAGTA	2400
	TCTAATAACA TTGTATTTAT TcmTGAAGGC ATGATCGGAG AACAAGGGGC TCCAGAAGAG	2460
50	ATGTTCAATC GTCCGAAAAC AGAAGAATTA AGACGTTTCT TAAATGTTAT AAATGAAGAA	2520
	TAATCAAATA GAACCACGTA TCATGTTTTA GTATGGCGAT GAAGCCATAT ACATGATGCG	2580
	TGGTTCTTTG TTATGTTGTC ATAATCTTGG AGCGATATTT TAACGACGTT TATGATTTAA	2640
55		

TTCTACATGT GCGTTAAAAC CTTTTTTGAA TTGTTGGACG CCATAGTCTT CTGATGACTC 2760  
 TGAAAAGTCA CCGGTAATAC CATAAAAATT ATAGCGATCA ATATGATGCG CTTTAGCAAA 2820  
 5 CTTAATCATT TCCCactGCA AATGGTAGGC ACCCATATAA GCATTATATT TAGGGTTTGA 2880  
 ACCACTAGAT AAGTAATAAA CTCATGCTC ATTGTAGATA AATAAAGCAG AAGCTAAGTT 2940  
 TAAGACTGCA CCATCTTGTT CAATTTGTTT TATTGTATTG TCGATTTTAC GCTTATTGCT 3000  
 10 ATTTAGCTGT TGTCTAGCT GTGTGCGTTT CGTTTTATTT TTCTTTGAAT TAGGACTTTC 3060  
 TTCCAATGCT TCTCAACAC CTGAGAGTTC AGCTGTTAAT TGTGTTGCT TTAATTGTAA 3120  
 15 CGTTTTTAAA TACTCGTTTA AATCAATATA CGCCAACTTT AACATGGCGT GGTCAATCGTA 3180  
 TAACTTTTGC ATTTCTTCAA AGTATGGTAA CTCACGGAAT TTGAAACCGT GCTTTTCCTC 3240  
 AGCCATATGG AATAAGTCGA AAAAAGTTTG CGTTTCATCA ATCGTTAAcG TTTTaGTTTT 3300  
 20 GACACCAaTa TCaTATGTtT tTtTAATATT ACGTCTCGtT TGATAATCCA TTTCTTTTAA 3360  
 AAGTTGGTCT TCAGTCTTAT CTTTTAAATC TAACACTGAC AGCCAACGGA TTTGGCTCAT 3420  
 TGAATCATAA CCTACAGGGA AACCTTGGTG TTTATAACCT AATTTATCCA TTGTTCTaAC 3480  
 25 AAATGCTCGG TTATCATAAG ATTTAACAAT TTCACCGTCT GCATTGCGTA AATTTTCAAT 3540  
 TAAATATGGA TCTACAAGGA CATATAAACA ATTTGTGTTT TTTAAATATG ACGTTAATGC 3600  
 TTTAAAGAAA AATGCTACTA ATGATTGATT TGTATAATCC ATCACTGGCC CGCGATGTGT 3660  
 30 ATAAAAATAT TTGAAAAATT TAAGTGTGCG TGCTTCTGTC AATAAGCATC CTGCAATCAC 3720  
 TTGACCATTG TCATCTTTAA CCCCTACrAG ATGCACATCG CCTTTTAAAT CAACTCTATG 3780  
 35 ATTGTAATGA ATAGCTGATT GTGTGTAATG TGAAAAATGC T 3821

## (2) INFORMATION FOR SEQ ID NO: 305:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1422 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

GCGACACCAT TGTTCGAGC AACTGTTAAT TTACCAGTTG ATGGTTTCGA CCGTGTGTT 60  
 50 GGTTCCTGATG GTGTAGTAGG GGTAGTTGTA GATTGCGTAC CCCATGGCGC CACTTTACCC 120  
 ATTTTATTATTA AATATTTTTC ATTAATTAAG TCATATAAATT GATCATAACT ATAATTATGA 180  
 CTTCTTAAAT ATCCATGTGG ATcGGCATGG TCAGTACCAC CTAAATATTT ACTTACAGCG 240

TATTGTAATT GTGTAGCTGC ATAGTCAGCA TAGTTATTCA TTGAACGTGC AAATGAAGCA 360  
 TAGTCGTGTG TGTGTACGAT TTCAACATTG ATGAATCTAG GGTTACCGAC TGCACCGACA 420  
 5 CCCCCAAGATA AGTAATCCGT TGGTGCTGTT TCGATTATAC GATCCCCATC AACAAATGCA 480  
 TGTACGAATG CGTTTTGATA GTTATTTTTC ATATAACTAA TTTCACCATT TATCGTCGAA 540  
 CGATCATTAG CTGTATCATG AACTACGATA CCTTCAGGAC GACCTACGCC GTTACGGTAT 600  
 10 GCGTATTTAG GGAAGTAAGA TGTATAATCT TCTTCAATTT TAGGTGCTTT TAAGTTATTT 660  
 TTACAAATGT AATCGTTAAT TGAAGAGTTT ACTTGTGGTT TATATTTTGG CAAACTCGTT 720  
 15 TTTGGTGTG CAGCAACTGA TCTTGGTTGT GCTGAAGCGC TAAAAGTAGT TACTTTAGGT 780  
 GTCGCTTCAG TTTTAGCTTT AGGTGCTGAT GTAGTTGCAG CTTTAGGTGC TCGCGTTTAA 840  
 TATTGCGTYT CAAGAGCTGC AGGTTTAGCA GCTGATTAA TTAATTCTGG ATTAATTGA 900  
 20 TTTTCTGAAT TATCATCTTC ATCATCAACT AACTATAAC CAGCATTGT AACATTAGTG 960  
 TTAGTTTTAG GTGCTGTAGT GCTTGTGAC TTTGCAACAG GCTGCGTATT ATTTGTAGTC 1020  
 GCTGATTGAT TAGCACGAGT GTCACCATT ACTTGTGCAG TATCAACTTT TTGACTTACT 1080  
 25 TGAGCATTGC CTGTTTTGTT ATTTGCTGTT TTTGGTTGGA CAATAGCAGG GTCTTGATAT 1140  
 ACTTGAGTGC CAGAAATGTT TTGCGTTGGA TTTTTTACCT CAGCTTTTGC TTGTTTCACTA 1200  
 GTTGCTTTAA CTTTATTACT ATCTAAAACG TTTTATTAG TAGTTTGATC TTGTGTCGTC 1260  
 30 TCAGCTGCTT GAACTTGATG TGCAGTGA CTGGAACCTA CAAGCGTTAA TGCAACCATT 1320  
 GAgGTAGTTT GTAATTGAAT TTTTCGCCA TTCTATTTAT TACTCCyAAC ATTTATTAAT 1380  
 35 TATtACTAAC ATTATAGTAC CTGTnTTATA TACCTGTGCG TA 1422

(2) INFORMATION FOR SEQ ID NO: 306:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6076 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

ACGCGTTGGT TATTTCAATC ATAATATTAC TCTGCAAATA CACAGCCCTG TAACCGAGTA 60  
 AATGGATAGA GATTGAACA AATGAAAACA ATCAACTAAT GGAAAGGATA AAATATTATG 120  
 50 CACAACAATA ATGAATTCAA CAAAAAAGT AAAGATTTTA TAGGTAGCGA TAAACGGATG 180  
 GCTTTAGTAA AGGGTTATGT TAACGAGTAT AAATTAACAA CTGTTTTAAG AGCTTTAAAT 240

	GTAGATAATC GAATTATTCC TAAAAATATA ACTCAAACA CAATCTTCAA ATTGAGTAAT	360
	TTAACCTTAG AGGTTAACCT CTATGAAAGA AACACAATTT ATCACGGAGA CTTTGCTATT	420
5	TACTATCCTG TAGAATCTGC TCTAATAATA GAAACAGATA CTCCTAGGCT AATTAATCAT	480
	ATAAATAACA ACCCAGTCCA AAAAATATTT ATAATAACAA CAAATGACTG GAGCTTTAAT	540
10	ACACTTGaAT TAGAAAAAAT AGTAGACGAA ACTATAATTT ATGATTTAAA ACAAGAAGAT	600
	CCTAAACAAT ATGAAATTTT ATATAAAAAT AAACACGGAA AATTGCCTTA TTGATGAATT	660
	AAAACGCAAA CTTAGAACAC CTTACCCCTT TTCCGTCCTT ATTGTAATTT TCCAACACT	720
15	CAAAATAAAA ACCACCCGTG CAAACAGGTG GTTTTATTAT ATATAATTAT TCCCACCTCAA	780
	TTGTGCTTGG TGGTTTTGAT GTAATGTCAT AGACTACGCG GTTGACGTGA TCTACTTCGT	840
	TTACGATACG ACTAGAAATC TTTTGTAAGA CTTCCCAATC GATGCGTGCG AAGTCACTTG	900
20	TCATACCGTC GATAGATGTT ACTGCACGAA TACCTACTGT GTGATCATAC GTACGGTAGT	960
	CTCCCAT AACCTACTGAT TGAATGTTTG GTAACACTGT GAAGTATTGC CAAATTTCTC	1020
	TTTCAAGACC TTCTTCTCTA ATCACTTGGC GTAAATCGC GTCTGATTCT CTAACGATTT	1080
25	CTAGTTTATC TTCAGTAATT TCTCCAAGTA CACGAATACC AAGACCAGGT CCTGGAATG	1140
	GTTGTCTCCA TACTAAATGT TCTGGAATAC CTAACCTCAAT ACCTAATTTA CGTACTTCAT	1200
30	CTTTAAACAA TGTATTGATT GGTTCGATTA ATTGCAATTC CATGTCTTCT GGTAATCCAC	1260
	CAACATTGTG GTGTGATTTG ATTGTTTGTG CTGTTTTAGT ACCTGATTCTG ATGACGTCTG	1320
	TATATAGTGT TCCTTGCGCA AGGAAGTCTA CACCTTTCAG TTTTGATGCT TCATCATCAA	1380
35	ATACGTATAC AAATTCATTA CCAATGATTT TACGTTTTTG TTCAGGATCT GAAACACCTT	1440
	TTAATTTATT CATAAAGCGA TCTTTGCGAT TAACACGAAT AATATTCATG TTGAAACCTT	1500
	CACCGAATTG CTCCATAACC ATGTCGCCTT CACCTTTACG AAGtAAGCCA TGGTCTACAA	1560
40	AGATACATGT TAGTTGATCA CCTATTGCTT TATGCAATAG TACAGCTACA ACAGATGAAT	1620
	CTACGCCGCC ACTCATCGCA CATAATACAC GACGGTCTCC TACGCGTTGA CGAATCTTTT	1680
	CAATTTGAT TTCGATAAAG TTTTCCATTG TCCATTGACC TCTACAATCA CAAACACGAC	1740
45	GGACAAAATT ATTTAATAAA TCATTACCAT ATTCTGTATG ACGTACTTCT GGATGGAATT	1800
	GAACACCATA AATGCGACGT TTCTTATCTT CGATTGCTGC ATAGTCTGTG CTTGGGCTAT	1860
	CAGCGATAAC TTCAAAGCCT TCTGGAATTT CAATAACTTT ATCAGAATGA CTCATCCAAA	1920
50	CAGTTTGTTC TGCTGGTAAG CCAGCGAATA ACTCATCTGA CTTGCGATTA ATGATTGCTT	1980
	TACCGTATTC ACGTTCATTG GCACGTTCAA CTTTACCACC TAATAATTTA GTAGTTAATT	2040

55

	TAAATGAACC TTCTTCATAA ACTGAATTTG GACCACCTGA TAAGATAATA CCTTTTGGAT	2160
	TCATTTTCTT AATTTCTTCA ATTGAAATTT CATGATCGTG TAATTCAC TAACGCCCA	2220
5	TTTCACGAAT TCGGCGTGTA ATTAATTGGT TGTATTGGCT ACCAAAGTCT AAGACAAGGA	2280
	TTAACTCTTG TTCTTTTGCC ATTTCCATAT TTGTCGTTCT CCTTTATCTT AATTAGAATG	2340
	AGTAGTTCGG TGATTTCTTC GTAATTTGAA TATTATGTGG ATGGCTTTCT GCTAAACCAG	2400
10	CAGGACCCAT ACGTGTAAT TGTGCTTCTT CGCGTAATTC TCTTAAATCG TGTGAACCAG	2460
	TATAACCCAT ACCAGCACGC ACACGCCCA TTAATTGGTA AATTGTATCT TGTAACGCAC	2520
15	CTTTATAAGC CGTACGTCCT TCGATACCTT CAGGAACAAA TTTCTTAGGC GCTTTGTCCT	2580
	CTTGGAAGTA ACGGTCGTTT GAACCTTTTT CCATCGCACC TAAAGAGCCC ATACCACGGT	2640
	ATACTTTATA TTGTCTACCT TGGAAAATTT CTGTTGCGCC TGGGCTTTCT TCAGTACCTG	2700
20	CTAATAAGCT ACCTAACATA ACCGCATGTC CACCAGCAGC TAATGCTTTA ATGATATCTC	2760
	CTGAGAATTT AATACCACCA TCAGCAATGA TAGCTTTACC ATGTTTGCGT GCTTCAGTtG	2820
	CACAATCATA AATTGCTGTA ATTTGTGGTA CACCAACACC TGCTACAACA CGCGTCGTAC	2880
25	AAATTGAACC TGGGCCAATA CCAACTTTAA CAATATCTGC ACCCGCTTCA AATAAATCTT	2940
	TTGTTGCTTC TGCAGTTGCT ACGTTACCTG CTACTAATGT GATTTCTGGG TAAGTCTTCT	3000
	TAATATGTTT CACTTGATCG ATAACACCTT TAGAGTGACC ATGTGCTGTA TCGATAACTA	3060
30	AGACATCCAC ACCTGCTTCG ACTAATTTTT GAGCACGAAT ATCAGTATCT TTTGAAATAC	3120
	CAATTGCTGC GGCTACAAGT AGACGACCAT GTTCATCTTT TGCTGCATTA GGAATTCGA	3180
35	TAACTTTTTC AATATCTTTA ATAGTAATAA GACCTTCTAG ACGTCCGTCT TTAAC TAATG	3240
	GTAACTTTTC AATCTTATGT TTTTGAGAA TTTTTCTGC TTCTTCAAGT GTTGTATTCA	3300
	CTGGAGCTGT AATTAAATtT TCTTGCGTCA TTACATCTAC AATTTTAATC GAGAAGTCTT	3360
40	CAATAAAACG TAAGTCACGG TTTGTAAAA TACCTACTAA GTTGCGATCT TCTTTATTAT	3420
	CAACAATTGG TACACCTGAA ATACGGTATT TACCCATTAA TGCTTCTGCT TCATAAACGC	3480
	TTTCTTCTGG CGTTAAGAAA AATGGGTTTG AAATGACACC ATTTTCTGAG CGTTTTACTT	3540
45	TTTGAAC TTC GTCCGCTTGT TCTTCAACGC CCATATTTTT ATGAATAACA CCTAAACCAC	3600
	CTTGACGAGC CATAGCAATC GCCATTTTAG ATTCAGTTAC AGTATCCATA CCAGCAGAAA	3660
	TAAC TGGAAT ATTTAATTTA ACTTTGTCTG ATAATGTAC GCTTAAATCA ACGTCTTTTCG	3720
50	GTAAAAATATC AGATTGTGCT GGAATTAATA ACACATCATC AAACGTTAAT GATTCTTTTG	3780
	CAAATTTACT TTCCACATT AAAACAGCC TCCATTTTTT AAATTAATTA GTTATATTAT	3840
55		

	GCAGAGATTG CGCCTAAAAC AATTCCGTTT TGAGTCAACC ATGCAAATTG TTCACCTAAA	3960
	CCTTTAAATG CTTGTGGTAC AGCGCTTATA CCAGTACCTA AtCCTACTGA TACAGCGATA	4020
5	ATTAATAAAT TGTTTTGATT TTAAAATCG ATATGTCCTA ATATACTAAC ACCATATGCC	4080
	ATTACCATGC CAAACATAGC TATCATCGCA CCGCCTAACA CAGGTAGCGG TATGATATTT	4140
10	GCTAATGCGC CAAGCTTAGG TATACAACCA CATATAAGTA ATAACACGAC CATGCCGTAT	4200
	ATAACATTGT TTTTCTTAGC GCCGGATAAA GAAACAAGTC CTACATTTTG CGAATAGGCT	4260
	GTATACGGAA ATGAATTGAA TATAGAACCT AAcACTATCG CTAGACCTTC CGCAGTATAA	4320
15	CCTTTACGAA AATCTTTTCT TTCTAACTTC TTACCGGTAA TTCACTTAA CGCATGATAG	4380
	ACACCTGTCG ACTCAATTAA ACTAACGATA GCTACAATAA AGAACACTAA CGTCGATGTC	4440
	ACATCAAAGC TAAATCCAGA GAATCTAAAC GGCCTGGGA TGCCTAACCA ACCGGCATGA	4500
20	TTGACTTGAT TAATATCGAC CATCCCAAGT AAGCCAGCAC CTATCGTTCC TAAAACGAGT	4560
	CCAATTAATA TGGCAATACT CTTAATAAAT CCAGTTGTGA ATCTTTGTAA AAGAAGAATA	4620
	ATGATTAATG TCATTAAACC TAACAAAATG TTCTTAACAT CTCCATAGTC CTTTGCACCT	4680
25	TGACCTCCAG CTAAGTAATT CATTGCTACT GGCATTAAAT TGATACCAAT GATAGTAACA	4740
	ACACTACCCG TTACTACTGG TGGGAAGAAT TTTACAAGAT GTGAAAAGAA AGGCGCGATG	4800
30	ATAATAACTA ATATCCCTGA TAAAAATAGC GAACCATAAA GTACATCTAT TCCTTTCGTT	4860
	TGACCAATTA AAATCATGGG CGCAACAGCC GTGAATGTAC ATCCAAGAAC GATTGGTAAT	4920
	CCTGTTCTCG TTACTTTATT GGCTTGTAAG AATGTGGCAA CCCCACACAT AAATATATCT	4980
35	ACTGTAACTA AGTAAGCGAT TTGTTTCAGG GTAAACTTCA AACTTGTACC AACAATGATT	5040
	GGAACCTAAGA TAGCACCTGC GTACATAGCT AAAAGATGTT GAACACTTAG GATTAAATTT	5100
	TTCAATTATC TTCTCCACC AATGTCACTT TGTTTCCTTC TAGTGAAGCA ACCTTGCAGA	5160
40	GAGAAGAAAC TGTTAAACCT GCTTCTTCTA AACGTTGATG CCCATTTTGG AAACTCTTTT	5220
	CAACAACAAT ACCAATACCA GCTGTCTTAG CATTCGCTTG CTGTGCGATA TCGTATAATC	5280
	CTAATGAAGC ATCACCATTT GCTAAAAAGT CATCGATGAT AAGTACAGTA TCTTCTTCTG	5340
45	ATAAAAACTC TTTTGAAACA ATGACCGTAC TTGTTTTATT TTTAGTAAAT GAATGAATAG	5400
	ATGTTTCATA ATAACCATCC GTCAAAGTGC TAGGTTTTGC TTTTTTCGCA AATAAACATG	5460
50	GCACATCAAA ATGCAGTGCA GCCATGATTG CAGGTGCGAT ACCGGAAGCT TCAATGGTTA	5520
	AGATTTTAGT AATCCCTTTA TCTTTAAATT GCTCGTAAAA AGTGCGACCA ACTTCATTCA	5580
55	TTAACTTTGC ATCAATTTGA TGATTTAAAA ATCCATCGAC TTTTAAAATC TTCTCATCAA	5640



ATTTGTGTGA AACATTTTGC TCTTAAATTG GTGCTAGATA CAAAAAATC CCCAACTAA 5760  
 ATAATAGTTT CAGGGTTTAT GAGTGAACGa ACATGCATAA CGAATTTGTC ATGCAATCAA 5820  
 5 TGTAAGAGAA GTTTCATCAA ATAACTGTG ACCATCATAT AAAATGATAT AAATCACCCA 5880  
 CCATGGTTAC AATTTAATGG CTGAAGCTAC TCCTAGTATT GTGTTGTTAC TCATAGTCAT 5940  
 GTCGTTCAAG GCAACATGGT AGAACTTCT AAAGCCATAT TCTTTAGATT ATATGAGTTT 6000  
 10 ATGTAAATTA TTTAACGATA ATAGCAAATT TTCGGCATT TTTCAATAAC TGCTTAGGTA 6060  
 ATCTTTTAAT AGTTTT 6076

15 (2) INFORMATION FOR SEQ ID NO: 307:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6136 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

25 ACCGCCGATT GATAGCTTTA CTGCTGCGAT AAAACTGCCT TCAACAATTG GTGCATCAAC 60  
 TTTTAACACA CGATGATTAC CATCATACAT TTCAATTGCC ATATCTACAT TCATTTCTGA 120  
 AGATCCAATA TCGTAAAAAC ACAATGCATC ATCCTCTAAT TTAGTCAAAA CTTCTTGGAT 180  
 30 GATATCAAAT GAAGTTCCAA TTGAACCATC TGGTAATCCC CCGATTGGTA TAATATCAAC 240  
 GTCACCTGCC ATTTGCTTTA ACAAAGATTT TGTACCACTT GcAATTTCTT TACTGTGGCT 300  
 AACAAGTATA ATTTTAGGCA TTATCTTCAT CTCCAATCAA AGCGTTTAAA ATATAAACCA 360  
 35 TACTTTGAGC ACCTGGATCA ATATAACCTT TTGATTCTTC TCCAAAATAT GCAGCTCTAC 420  
 CTTTCGTTGC TACCATATCT TTTGTATTAT CTGCTAATTG CTGTAAATCA TTGAATGTTA 480  
 40 AAGTTTCACC ATTTTAAAGC TTCTCTGCTG CTCGCGCTAC TACATCATAC ATTGTCTTTT 540  
 CATTTAAAGT AACTTTACCA CGTGATTCAA CCGCTTCGGC AAATGCCTGA ATTAGTGTA 600  
 TGAAATCTTG ATTATCCATA TCATCTTTGG TGAATGCAGA CATTTTAACA AAGCTAAAGC 660  
 45 CATACAGTGG TCCTGATGCA CCCCCAACAT TTGACATCAA TGCCATACCA GTTGATTGTA 720  
 ACAATGATTG CATTGAGCTA TCATCAAGTT TGTCTTTAAG ACTACTAAAC CCACGAACCA 780  
 TGTTAACCCC GTGGTCACCA TCACCAATTG CTCGATCTAA TTCAGTTAAT TCAGATTCAT 840  
 50 GTTTTTTAAA CGTTTCTTCT AAATTTAATA AACGTGCTTT CATATCATTC ACTTTCATTT 900  
 GTGCCACCTC ATAAATGTAT ATTTATTCAT ATTCACCTTCT TATTTAAAGT ATTGACTTGT 960

55

	ACCTTGCA TG TCTAAAGATG TCATATAATC ACCAACAAAC CATTTAGCAA CATTAACCGT	1080
	TCTTGCA GCT AAATTTTGTT GAATATATTT AGTTACGATA TTTAATTCAG ATAACGGCGT	1140
5	ACCACCCATA CCATTACCA TTAATATGAC ATCATTGGCA GTAACCTCTT TATACAATTC	1200
	GTCTAACAAT GTTCCAACAA TATGATCAAT ATCCTTTACT TCTTCCCTAT GAATACCTTT	1260
10	TTCACCATGT ATACCAATAC CGATTTCAT TTTGTCGTCT TCAATATCAA AGCCATATTT	1320
	TCCAGTAGTT GGAACAAGCG GTGGCTCAAT TGCCATACCA ATACTTTTAA TTTCAGGTAA	1380
	CAACGCTTCT ACACGCGATT TTATCTCTGT TAATGAATAA CCTTTTTCAG CAAGATAACC	1440
15	GGCAAGCTTA TGAACAAACA CTGTTCTGCG AACACCACGA CGTTGTACTT CGTTTGTGAC	1500
	AGCAATGTCG TCACGAACAA TAACAGTTTG AACATTTATA CCTTCCATTT CTGCAAGCTC	1560
	TTGTGCCATT TCGAAATTCA TCACGTCACC TGCATAGTTT TTTACAAC TAATACACC	1620
20	ATCACCAGTA TCTACTGCTT TAATAGCTTC TAATATTTTA TCAGGTGTAG GTGATGTAAA	1680
	TACTTCGCCA CAAACCGCTG CATCTAGCAT ACCTTCTGCA ACAAAAACCGG CATGCGCAGG	1740
	TTCATGTCCG CTTCCACCTC CAGAGACTAT TGCTACACCA TGTCTTTTCT TAGCTTTTTT	1800
25	TACAATAACT GTATTAGCAA TCAGATCTAA CTCTGGGTGC GCAATTAACA ATCCTTCAAG	1860
	CATATCAGTT AAAAATGTTT CTTTTTTATT GATTAACTTT TTCATCATGT TGTACCTCCT	1920
	TGGTATTATC AGTCATAGTA TAAACAGAA TAAATGAATG CGCTATCATA AAGAATTAAC	1980
30	CTGATACCGT TATCAAATAT ATTTTGTGATA AGATCCTCTT GATAACTTAA TGGTTCATTA	2040
	TTGAAAAAAT AAATAGTCTT GAGTGGCTCA TTAAATAAAA AATGCTATGA AAAGCCTTAT	2100
35	AATAAAGTGC CTTTCATAGC AATAAGTTGT GTCCATTGAC ACTATACATT TTTCGTTTTG	2160
	TACATTAAAT ATAAGAAATA CGGTGCACCA ATAATTGCTA CGACAATACC TGCTGGAACC	2220
	CCACCTGGTT GTAATACAAT TTTGCCAATT GTATCAGCTA TAACAAGTAA ACATGCCCCCT	2280
40	ACTAAAATGG CAATTGGTAA AAACAACTGG TGACGTGGTC CAACGATACG TTTGGCAATA	2340
	TGCGGACCCA TTAATCCGAT AAACGAAATT GAACCTGCTA CTGCTACAGC AGCAGATGAT	2400
	AACATCACTG CGATAAAGAA TAATATTAAA CGTTCTCTGC TTAACCTTAC ACCTAGACCT	2460
45	CGTGCAATAT TATACCCCGT ATGAATAATA TTTAGTGAT TCGATTTAAA TAGTAAGTAA	2520
	GGAATAATAA TCAACACCCA CGGTAAAAAT GCAATGACAA ATGGCCATTC GTCACCCCAA	2580
	ATATTACCTG CAAACCAAGC AGCGATGAAA TCAGATTGCT TATCATCAAA TTTTGACATA	2640
50	ATTGTAATTG AGCCACCATA TAATGCTGTT TGTAACCTA CACCTATTAA TACCATACTC	2700
	GCAGGTGTAA CACCTTCATT TTTATTAAAA CTGAATAATA AAATAATCaA TGCAGTGgTG	2760

55

	CCAATTGCAA TAAATAATGC GATTGCAAAT CCGCCACtGC GTTAATACCT AATATACCTG	2880
	GTTCAGCTAT TGGATTTTTT GTGACACTTT GCACTATTGC ACCACTAATA CTAAGCtGCG	2940
5	CCAGCCAAAA TAGTAATCAT CATCCGAGGT AACCTGAAAT CTAATAAGAT TAACTCATCA	3000
	ATGGCATCAC CTTGTCCAAT TAAAGTTTGT AAAAATCTTT CAACAGGTAT GTTGTATTCA	3060
	CCTGAGGTAA TACTCCAAGT ACAACCTAGA AATAGTAGAA TGCTAAAAAC AGCCAGTGCT	3120
10	ATCAATTGTC TGCCTTTATT ATTTGAACTA ATCATATTGA GCGTCCTCCT TTTTAACTA	3180
	AATATAAAAA GTAAGGAACA CCGATAAATG AAATGATTGC ACCAACAGGC GCTTCTCCTA	3240
15	AATATCGTGC TATCACATCG GCAACAAGCA CGAGTATCCC ACCTAACAAG GCTGTTAATG	3300
	GTAGAATTTT AGCATAATCA GTTCCAATTA AAAATCTTGC TATATGAGGT ACCATCAAAC	3360
	CTACAAATGC AACTTGTCCA GCGATAGCAA CTGCAATACC TGCTAGAATC ATAGCAATAA	3420
20	TTAAACATAT GCCTCTGATC ATTGTTACAT TTTGACCTAA ACCTTTAGCT AATGATTAC	3480
	CAAGATTTAA AATGGTAAGT TGTTTACTAA TTGTTAATAT AATGAATAAC GCAATACCAA	3540
	TTAATGGAAT TGCCCACTTA AGGTGTGACC ATGTTGTGCC TGAAACGCCT CCAGCAGTCC	3600
25	AAAATGTTAC TGTTTGATT AGTCTAAAAG CTAATGCAAT ACCTTGACTT AGCGCTGTTA	3660
	ACATAGCACT TACTGCTGCA CCCGCTAAAA TAATACGCAT CGGATTAAAT CCATCACGTC	3720
	TAGATCGGCC TATCATTAAAT ACAATAGCAC CTCCTAGAAT AGCACCTAAA AATCCAGCAA	3780
30	ACATCAATAT TAAAAATGAA GTGTTTGTA AAACCTGCATA TGTTAATGCT AAAGCAAATG	3840
	AAGCACCTGA ATTTAAACCT ATGAGCGCCG GATCAGCAAG ACCATTACGA GTAACACCTT	3900
35	GTATAATCGC ACCAGAAACT GCAAGCGCCA TACCTACAAT TACTGCTGCT ATATTTCTGG	3960
	GAATCCTAAT CTCATTGATG ATGTTTTGCT GTTGATTGCT AGGATTATAA TAAAAATAG	4020
	CCTCTATAAT TGTAGAGGCT TGAATTTTGG CGTCACCTAT TAATGTAGAA ATAAATAGTG	4080
40	TGATTAGTAG TATCATACTT AAACCTATAA TATAGGATAA AAACCTCAAT GCGTTGGGT	4140
	TCTCTCTATT TGTCATGTTA ATTGTCCTTT TTATCATATT AACTTACTTA ATTAAGAATA	4200
	AGCTCTGCGA CATAAGTCAT AAGTTACCAG TAAAGGTTTT CCAGTTTTAG GATCTTTACT	4260
45	TAAACAACA TCAATATTAA AAACCTTTTC TAATATTTCC TGTGTTAATA CGTCTTCTGT	4320
	TGAACCTGTA GCGATGATAT CCCCTTCTTT CATCGCAATA AGATGATCTG AGAAACGAAT	4380
	CGCTTGTTTG ATATCATGAA GAACCATGAC AATTGTACAA CCTTGTTCTT GATTTAGCTT	4440
50	CTGAACATA TCTAGTATTT CTAATTGATG ACAGATATCT AAATATGTTG TTGGTTCGTC	4500
	TAAAAAGATA ATATCAGTTC TTTGTGCTAA TGCCATTGCA ATCCAAACAC GTTGTCTTTG	4560
55		

TGCCCAATCA ATtCTTTCT TAtCCTCAGC AgTTAATCTA CCAAATCCTT TTTGATGTGG 4680  
 AAAACGACCA TATGAAACTA ATTCCCAAC AGTTAAGCCA TCTGCTACTT CaGGTGaTTG 4740  
 5 aGGTAAaTG GcAtTTTTt TGcAATCyCy TTCGTAGAt GTGtATGAAT ATTTTCACcA 4800  
 TCTAAAAATA CTTCGCCTTC TTAACTGCC AATAAACGTG ACAATGCCTT TAGCAAAGTA 4860  
 GATTTCCCGC AGCCGTTAGG ACCAATGATT GACGTCACCT TGCCATCTGG TATTTCAACA 4920  
 10 TCTAATTTAT TTATAATCGT GTTATCCCCG TAACCAATTT TAACTTGTTG TCCATGCAAA 4980  
 CGATTcATAA TTTCCCTACT TTCAATAAAA TTCTTTCTGT TTATAAAAAA TAATTTCTAT 5040  
 15 TTTTAAATTA TCAATTTTCA AAGACATCCC AATTGATAAT GATTATCATG AACATCATTa 5100  
 TAACATTTTT CAATCTTATT GACTAACATT ACTTTTTAAa TTGGATAGCT CGATTTGTCA 5160  
 TGTCTTGtAT ATTACTTTTA TAAAATAAAA AACGCCcACA GATAAGTCTT CATAGTTCAA 5220  
 20 AAACCTTGtCC GTGGACTTCT ATTTAAGTAT GTGTGCTCAT ACCATTTATt TATTCATCTG 5280  
 CAAGAAAGCC ATTACCATAG ACATCTCTTA CATCATGAAT TACGAGGAAT GCATCTTTAT 5340  
 CGATTTGTTT AATTAATCGC TTTGCTTTTG AAACCTTGtGT TTTAGAAATA ACAACGTATA 5400  
 25 AGACATCTTT TTCTTCACGC GTATAATAGC CATGTCCGTT TAAATGGTT AAACCTCTTC 5460  
 CAATTTGCTC GTCTATTGCT TTGGCAAGTT TGTCGGGATT AGTTGAAATA ATCGTCATAG 5520  
 CTTTTTTAGT GTTTAAACCT TCTATGACAT ATTCCATCAC TTTTGTTcCT ATATAAAGTG 5580  
 30 ATATTACTGT TACTAATACT TTATCAAGTG GAATAACTGT AAGTGAAATT GCAACAACGA 5640  
 TCATATCGAA GAAAAGCAAA GCATATGGCG TGCTTACATC GAGGTATTTc GTTGCAATTC 5700  
 35 TCGCCAAAAT TGTTGTACCT GCTGTTGTAC CGCCTGCAAG GATAATTACT CCGATTcCTA 5760  
 GTCCAACGCT TACACCACCA AAAATGGCAT TCACAATGCT GTTTCCAGTT TCTACTTGCC 5820  
 ATGATTCTGT TAAACTCAAA AATATTGAAA TAAGAATTGT TACAAGAATA GTTAAGTACA 5880  
 40 TACTTCTCTT ACTCAAAAAt TTATAAcCTA TGGCAATCAA TACTGCGTTG ACCAAGAAGT 5940  
 TAGTGATGGC TGGTGAAATA TGAAACGCAT AATATAAAAT AATTGCTAAA CCTGTAACCC 6000  
 CGCCTTCACC TAAGTTACCA GAAATaATAA ATGCATTTAC ACCTGCAGCA AAGATAAATG 6060  
 45 AACCTAAGAC AACTAGTATT AAATCTTTAA CCGTTTTATT CACGAAACCA TCCCCTTTAT 6120  
 ATATTTATTA GACTAT 6136

(2) INFORMATION FOR SEQ ID NO: 308:

50

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2576 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

5	GATATCGTAs	CTACTGAAAA	GTCATCACCA	CCATGGAATG	ATTTCTTTAA	ACGTTTTAGC	60
	TTCTATGCAA	TTGCAATTCA	ATACTTTGTT	GTACAATTTA	TCATTACATT	ATTCTTAATT	120
	TGGTTACCGA	CGTATTTAAC	AGAAGTATTC	CACGTTAACT	TTAAAGAAAT	GAGCATTAGT	180
10	TCATTACCTT	GGTTATTAAT	GTTCTTCTTA	ATCTTATCAG	CAGGTGCAAT	TTCTGACCGT	240
	GTATTAGGAT	TAGGTCGTTT	AAAATTCGTA	GCTAGAGGTG	TAATTGCAAT	TGCAGGATTT	300
15	ATTGTGTTTG	CAGTTTCAAT	TATCTTTGCT	GTACGCACAG	GAAATTTATA	TGTAAGTATT	360
	TTCTGGTTAT	CACTAGGTCT	TGGTGGTATC	GGTATTTCAA	TGGGTATGAG	TTGGGCTGCA	420
	GCAACTGACT	TAGGACGTAA	CTTCTCTGGT	ACAGTATCAG	GGTGGATGAA	CTTATGGGGT	480
20	AATATAGGTG	CATTAATCAG	TCCGCTATTA	GCAGGTCTAT	TCGTAGAACA	TTTGGGTGG	540
	ACAATGACAT	TCCAATTGTT	AATCGTTCCA	GCAGTAATCG	CTGTGATTAT	GTGGTTCTAT	600
	GTGAAACCAG	ATCAACCTTT	AATTGTTAGT	GATGATAAAG	CAATAGAAAA	ATAATTTAAA	660
25	CAAGCAGTAA	GCTTTCACAT	AGTTGGGGCT	TATTGCTTTT	TTTGCGTTGA	AATTGAAACT	720
	TTTTAAACA	GATATGGTTT	AAGATGAAAA	TGAAGTTATT	GAAATGATAT	ATGTAAAGAA	780
	ATAAGGTTTT	AAACATTAG	TCAGGTAACG	CTTGTA AAAAG	TACATATAAA	TTTTAACTAG	840
30	CGCAAAGGTG	GGCGACCAAA	GtTcaACGAT	GTAAATAAC	aTTAGrAATT	AATTTTAATT	900
	GGACTTTAAA	AGTTTTTAAA	TTTAGATAAT	TGAGCATAAG	GTGTTATAAT	GACATATGTT	960
35	GCGTAATTAA	AATTTATAGC	AACAAATTCA	TTTTAACTAT	GCTAATAAAA	AGATTATGGA	1020
	AATATTTTGA	CAAGGAAAGG	AGAAGTCGAA	ATGACATCTT	TTTGACATCA	CTCATAAAAA	1080
	TCAATCGACT	TAACCTAGAC	TTTTATAAAG	GTGTAAGACA	GGGACTGTTA	ATGATTATTC	1140
40	CTGCAATAAT	CGGTTACTTA	TGTGGTAATT	TCCAATTTGG	ATTATTAGTT	GCAACCGGAA	1200
	CACTAGCCCA	TATTTATGTT	TTTAAAGGTC	CGTCGCGATC	TAAGCTGCGA	ACTGTAATAA	1260
	TTTGTAATTT	AGCGTTTGCA	ATATGTATGA	TGCTTGGTAC	GCTAACAGCC	AAAACGCCAC	1320
45	TCGTTTTTGG	AATGACATTA	TTAATTGTTA	CGGTTATACC	ATTTTATATA	TTTACTGCCT	1380
	TAAAAATAGC	TGGACCGTCA	TCGACATTCT	TCATTGTGAC	ATTCAGTCTA	CCCATTAACT	1440
50	TACCTATAGC	TCCCGAAGAA	GCATTATATA	GAGGCTTTGC	GATTTTAGTA	GGCGGTATAC	1500
	TTGCCACTAT	GATGGTGTTA	ATCACGATCG	TATTTTCTAA	AAACAAAGCT	GAAGAACAAG	1560
	CAATTCAAAA	TGATTTTAAA	CTCATATCTA	AGTTGTTACA	CACTTATAAT	GATAAATCTG	1620

55

TCACTTCTAC TTCAAGTAAC GATAAATTAA GTAGACGTTT CCAAAAATTA TTATTATTAC 1740  
 ACACATCTGC CCAAGGGATT TATTCTGAAC TGTTAGAGTT GAACGCTAAA CAAATTGAC 1800  
 5 CATTGCCAGA TGAGTTAATT GAAATGATGG ATCATATCAT TGCACAAC TA GATAATAGTG 1860  
 AGGAAAATGT AAGATATTGG CGAAAAGAAG TGACAGTAAC AGAGGAATTT CAAAATTTAT 1920  
 TCAACCATAT ATTGAAAATT GATGAAATGG TGCATGCAAA TGAAGCGCGT ATTGCGTATG 1980  
 10 AAGCAGACAT GCGAAAACCT TTATATAGTA AACGCATTTA TCaAAATTTA ACaTTAGACT 2040  
 CtAtkGTTTT TAGAAATACA TTGAGATATA CAGCGATTAT GATGATAGCG ATATTTATTG 2100  
 CGTTAATGTT TGATTTTGAA AAAGCATACT GGATACCGTT ATCTGCACAT ACAATATTAC 2160  
 15 TAGGAACATC AACTATACAT GCAATCGAGA GAGGTATGGC ACGAGGTTTA GGTACTATTT 2220  
 TAGGTGTGTT AGTACTTTCA GTCATATTGT TGTTTTCAAT ACCAACACCT GTTGCGATAA 2280  
 20 TTTTAATGGG CATTGCAGCA TTGTTTACTG AAGCATTGGT GGGAGCAAAT TATGCGATTG 2340  
 CAGTAGTTTT TATTACAATA CAAGTTATTT TAATGAACGG ATTAGCATCA CAGAATTTAA 2400  
 CAATTAACAT TCGGTTTCCA AGAGTTATTG ACGTTGCAAT GGGTATTGTG ATTGCAATCA 2460  
 25 TAGGTTTATT TGTCTTGGA CAACGTACCG CATCCGCATT GCTTCCTAAT GTAATGGCTG 2520  
 AAGTTGTTTCG TAAAGAAGCA ACGCTCTTTC ATTATTTATT TTCTGAAAAT CAATAT 2576

(2) INFORMATION FOR SEQ ID NO: 309:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 668 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

40 CAAGCACATT ACGATGGTCT AATACTTTC TAATAATTTT TTCTTGTCTT GGTGAAACG 60  
 TTTCATACCC AAAGTAATGC GATAATGTTT GTTGATCAT AAATTGACCC CTTATTGTTG 120  
 TTCTTTAATT TCTTCTAACT CACTCCATCT TGTGATGTCT AAATCATATT GAATTTCAAG 180  
 45 TTGTTCTTTT TCTTCGTTTA ATTCTTTAAT TTCCCATAA TCTGCACTTG CCTCAATCAT 240  
 GAGCACATCA ATTTCTTCCA TTCTTACTTC CGCTTGTTCT ATGCGTTTCA TCAATTGTTC 300  
 ATATTCTAAT TTTCTTTTAT ATGATAAACC ATTTTCTTCA CGTACAGTTG TAGAAGATT 360  
 50 AGATTGTTGC TTCAATGTGG ATTTATTTTT ATCTAATGAT TTTTATAAC TTTCATAATC 420  
 TTCAAAAGTT CCGATAATCT TTCCATCTG ACCATCATGA ATAAACCAAT ATGACTGTGC 480

55

AATATAATCT TCAAGTATTG TTAAAGTCTC AGTATCTAAA TCATTGTGCG GTTCATCTAA 600  
 CAACAGAACA TTTGGCTGGT GTACGAGTAG ACGTAATAAA TACAAACGCT TTTGCTCTCC 660  
 ACCAGATA 668

(2) INFORMATION FOR SEQ ID NO: 310:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12173 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

CATCAGTTTA TTTGAAAGG CAATGCGATC ATTTTCATGTA TTTATGTTGT TTGAAACATC 60  
 GTTAGATAAC AATAGTGATA TTGCACATTT TAAGCTGAAG ATGGAAAAAT ATTCGATAAT 120  
 CAAATAAACA ATGAATTTTA GAAGGTACAA TGACGTTTAC TAATTTAAAT ATAGCTGAAT 180  
 GTGTTGGTGA GTGATGTTCA CTATAGATTT ATATTAATAT ACAAAGACAA AGGTTGTAA 240  
 TTTTATTAA GCGTTAGGTT GAATGTATGA GAATTTTAGA TTTATAATAG AAGATAGAAA 300  
 CGAAAATTTT TCTTAAAGC AGTAATGTTG ACTCAAATA AGCTATAATA ATGACACTTA 360  
 TTAAATTGAT TAACATTTGC TAATAAATAT CAATATAGAA TATAACTTTC CAATAATGAC 420  
 TGAGAAAATC GAAATGTCAG TCTCGAATCA TATAATTAGA AAATTGATTA TTTTCTGTCA 480  
 ATTTAGGGTT GAACTATACA TATGATATTG TTAGAATATT TTTTAACATT ATATTTTATT 540  
 GCTTTAAAGT GGAATATACT TGAAATAATT AGTAGAGGTG AGTAAGGATG AGTAATAAAT 600  
 TAGAATCATA CAGAAGTGAG ATTGTATCAC TGAATCATCA AATTTTAGAC TTATTATCTA 660  
 AACGTGGTGA ACTAGCACAA AAAATTGGGG AAGAAAAATT AAAACAAGGT ACACGTATTT 720  
 ATGATCCACA ACGTGAAAAA GAAATGCTTA ACGACTTAAT CGATAGTAAC AAAGGACCAT 780  
 TCAACGATAA TACTATTAAG CAATTATTTA AAGAAATTTT CAAAGCCTCT ACAGATTTAC 840  
 AAAATCTGA AAATGAAAAA CATTTATATG TATCACGTAA GTTGAAACCT GAAGATACGA 900  
 TTGTAACATT TGATAATGGG GGCATTATAG GAGACGGCAA TAAATCATTT GTATTTGGGC 960  
 CATGTTTCAGT TGAATCATTT GAACAAGTTG AAGCTGTTGC TAAAACTTA CATGCTAAAG 1020  
 GTGAAAAATT TATTCGTGGC GGTGCATTTA AACCACGTAC ATCACCATAT GATTTCCAAG 1080  
 GCCTAGGTGT TGAAGGACTT AAAATACTTA AACAGATTAA AGATAAATAT GATTTAAATG 1140  
 TTGTCAGCGA AATCGTAAAT CCAATGATT TTGAAGTGGC TGATGAGTAT TTAGACGTAT 1200

	AAAAGCCTAT TCTATTAAAA CGTGGTTTAT CTGCTACAAT CGAAGAGTTT GTTTATGCAG	1320
	CTGAATACAT TGCTTCACAA GGTAATCAAA ACATTATTTT ATGTGAACGT GGAATCCGAA	1380
5	CTTATGAAAA GCGGACACGT AACACTTTAG ATATTTTCAGC AGTACCAATT TTAAAACAAG	1440
	GTACACACTT ACCAGTCATG GTAGATGTTA CGCATAGTAC AGGTCGTAAA GATATCATGT	1500
	TACCAACTGC GAARsAgCAT TAGCAGTTGG TGCTGATGGA GTTATGGCTG AGGTGCATCC	1560
10	AGATCCATCT GTTGCACTTA GTGATGCGGG TCAACAAATG GATTTAGATG AATTCCAAGC	1620
	ATTTTATGAT GAATTAAAGC CTTTAGCTGA TTTATATAAC GCTAAAAAGT TAAAATAATA	1680
15	TTCCAAGGAA ACTATAGACT ACTTAACTAA TATGTCATGT TGAAGTAGAA TATTATCTTT	1740
	GAATCGACAA TTTTAACTT ACAGCCATTC TAAGAGTATA TTACTTTTAG AGTGGCTATT	1800
	ATTTTTTGTA TAGAAATAAA GGTATACTGC ACTTAACGAT TGTATAATA CTTCGACACT	1860
20	TGTTCAATTT CACAATTATT AAAGATTATG ACTGATAGCA GTAATTAAAA TTATAACTAT	1920
	GAATTATCTG TAAAATATAA TAGATTCACA CATTTGTTGC TGAAATGTGA ACATTTTTCA	1980
	ACAAATGCAA TTGATATTTG AAAAGGCTTT CTCAAACAT TACAATTAAA AATGAAAAAA	2040
25	GTATATATAA AATTAAAATA TATCGTTCGT TATCATTTAG CGTTTGTTTT TATTTCAAGC	2100
	TTTTCGCTAA ATTTTTCCAA ACAAAAATAT GTTACTGTAA ATTAAAATAT GGTAACTAT	2160
	GAAAATGAAA TGAAAACATG TTATTATAAT GAATAAACG TTTACAAGGA GGAAATTATG	2220
30	ACAGTTACTA TATATGATGT AGCAAGAGAA GCGCGTGTCT CTATGGCCAC AGTGTGCGGT	2280
	GTGTGTAATG GGAACCAAAA TGTTAAAGCA GAAACTAAAA ATAAAGTTAA CGAAGTCATT	2340
35	AAGCGTTTGA ATTATCGTCC AAATGCTGTT GCTAGAGGTT TAGCTAGTAA AAAGACAACA	2400
	ACAGTAGGTG TGATCATTCC AGATATATCT AATATCTATT ATTCACAACT TGCTCGTGGA	2460
	CTTGAAAGATA TTGCAACAAT GTATAAATAT CACTCAATTA TTTCAAATTC AGATAACGAT	2520
40	CCTGAAAAGG AAAAAGAAAT TTTTAATAAC TTATTAAAGTA AACAGGTTGA TGGTATTATT	2580
	TTCTTGGTG GTACAATTAC TGAAGAAATG AAAGAATTGA TAAATCAATC ATCTGTACCT	2640
	GTAGTAGTAT CAGGAACAAA TGGTAAGGAT GCACATATAG CATCAGTTAA TATTGATTTT	2700
45	ACTGAAGCTG CGAAAGAAAT TACGGGAGAA TTAATTGAAA AAGGCGCTAA ATCATTGCT	2760
	TTAGTAGGTG GAGAACATTC TAAAAAGCT CAAGAAGATG TTTTAGAAGG TTAACTGAA	2820
	GTGTAAATA AAAATGGCCT TCAATTAGGT GATACATTGA ATTGTTCTGG TGCTGAAAGT	2880
50	TATAAAGAAG GCGTAAAGC TTTTGCCAAA ATGAAAGGCA ATTTGCCAGA TGCCATTTTA	2940
	TGTATCAGCG ACGAAGAAGC AATTGGTATT ATGCATAGTG CAATGGATGC TGGTATTAAA	3000
55		



## EP 0 786 519 A2

	CCACAAC TTT CTAGTGTTAT TCAACCATTA TATGATATCG GTGCAGTAGG GATGCGCTTA	3120
	TTAACAAAAAT ATATGAACGA TGAAAAGATA GAAGAACCAA ATGTAGTTTT ACCTCACAGA	3180
5	ATTGAATACC GAGGAACTAC AAAATAAATT CACAAAATTA GGCATTTCATC TAACGACCCA	3240
	AATTATATGG GTGTTGGAAG AATGCCTTTT ATTATCTTT TAAAATCGTT GCAGATTAGG	3300
	TTACTTATTG ACGAGTAGAT TCGTACCAAC TCGCTATATG TAAAGCTAAT TTTTATTTT	3360
10	TTTCACTAAT TTCTTTTGTG CGGGGGACAT AGGTATAATC ATTTAAACGA TCTTCCCATC	3420
	TTTTAGGTAA TAATTCAGAT GAATAATGTT TCCATTATT AATCCATTCT AACGGTAAAT	3480
15	AACCACTTTG AATTGGTTGA TCAATTAAAC TTAAGAATAC ATGACTCCAT GCACGTGGTA	3540
	CGACTCTCCA AATATTGTAG CCTCCGCCAC CAAACATAAT TACCTTTCCA TTCGTATAAG	3600
	AATCAGCTAA ATATTTTACA AAATATGGAA TTTCATATAA TGAATGTAAC GTACAATTTA	3660
20	GATGAGTTAG TGGATCACGA TAATGTATAT CGACACCATT TACGCTTAGA ATAATATCAG	3720
	GTTTAAAACT CTTTACGACA GGCTCAACTG TTAATTAAAA AACTCCAAA AATGATGCAT	3780
	CTTCTGTATA CGGTTCAAGT GGGACATTTA CAGTGTGTCC ATAGCCGATA TCTTCACCGC	3840
25	GCTCAGTATA GTGACCAGAG CCTGGGAAAA GAAATTTTCC GGTTTCATGG ATAGAATAAG	3900
	TAGTAACATG GTTATCGGCA TAGAACTCC ATTGTGTACC ATCTCCATGA TGTGCATCGG	3960
	TATCTATGAT TAAAACGCGT TGATTGTATT CTTTAGCTAA GTATTGTGCG GTAATTGCAA	4020
30	TATCATTGTA TATACAAAAA CCACTTGCTC GACCAGGTTG AGCGTGATGC AAACCACCAC	4080
	CTAAGTGACA ACCATTTAAT ACTTTGCCTG ACATAATAAG ATCTGCTAAA GTTAAAGCGC	4140
35	CTCCAACAAT TGTGGCACTA TGGCGGTGCA TATGCTTAAA TTGACCATT CTTTCATCAT	4200
	TTAATCCATA TTTCTTAGCC TCATCTTCAC TGATAATGCC ATGTGAAGCA TGCTTAATAG	4260
	CTTCGACGTA ATCATATTTA TGAATTAAAC TTAATTGCTC ATCTGTTGCA ATTCTAGGTT	4320
40	GTACTATTG TTCTGGAGAC AATAAATTTG CATTCAAAAG TAGCTCTGTT GTTAATTTTA	4380
	AACGCATTTG ATTGAAGGGA TGTTGGTCAT GAAATCGATA TTGTAATAAC TTATCTGAAT	4440
	AAACATATGC AGTTTTTGAT GAATGTTGTT GCATATAATC CCTCCGATAT TCCAAAAATT	4500
45	AAAAGAAAAA CCGATTCATA TAACGAATAT CATCAAACGC TTGTTGCTGT TCTAATGTAA	4560
	TGTTTTTGCC AATTCTTGCC ATTAAACAAT TAGCTGGATG ACTTGTTATT TCTGGATCAT	4620
	CTGTAGCGAA TATTTCAAGT CCACCAGTTG CCATTAAACG CTGCATTAAAT TTTTATAGT	4680
50	CAAATACATC TAACTTTGAA TTTTTTAAAT CCCAATGCCA GTAATATTCT GTAGTTATAA	4740
	CGATATAATT CTCGAATTCT GGTGTAGAAA GGCTAAGTTG TATCAGCTTT TCTGCAAGTT	4800

55

TACCTGTAGA CCAGCGCTCA ATTTTCATCAG GATAGTGGAA AGTGACATAA CCCACAATGA 4920  
 GTTGATCTTG GCGAGCAACA TAAATTCTAC CTTCTGGTAA TGTGTGAATT TCTAACAAAG 4980  
 5 CTTTATACTG ATCTTCAGCA TCTCTAAATG CGGTAAATG CGCATCGAAA GTAAGCGCTT 5040  
 TCAAATCTTC GTGTGTTAAA GGACCTTCAA TAACAAATTG CTTGTCATGA ATGTAATAAT 5100  
 CTTCGGATTG ATACGTCTTT AAATGATTCA TATTTTCAAC TCCTCAATCG ACGTCGTGAT 5160  
 10 TGTATTAAAT TCATTATATA GAAaATTTAC AATAATTAAT ACTAGAAAAA GGATAAAAGT 5220  
 AAAAAATTTG AATAATTAGA AATGTTATGT ATAATATTGA GAAAGAAAGC GTTTTCACAT 5280  
 15 AACAAAGGGG GAGTTTCAAA TGAAAGTCGA AGTTTATAAA GGAGCGCAAG GTAAACATAA 5340  
 CCTTAAAGAT TATGAAGAAA CATATAATAC TTTTGATTGG AAAGACGTAG AACAGCATT 5400  
 TTCTTGAGT GAAACTGGAA AAATGAACAT GGCATATGAA TGCATAGATC GCCATGTAGA 5460  
 20 TCAAGGATTA GGGGATAAAA TAGCGTTAAA TTACAAAGAT GAGCACAGAA AAGAATCGTA 5520  
 TACTTATAAA GATATGCAAC GGTATCTAA TAAAGCAGCG AATGTTTGT CTGAACATGC 5580  
 AGAAGTTGAC AAAGGTGACA GAGTATTTAT ATTTATGTCG CGTACACCTG AACTATATTT 5640  
 25 TCGTTGTTA GGTGTTTTAA AAATTGGTGC AATTGTTGGG CCGTTATTG AAGCATTAT 5700  
 GGAAAGGCA GTTGCGGATA GATTAGAGAA CAGTGAAGCT AAAGTGTAA TTAATAATAA 5760  
 GGCATTGTTA CCTCGAGTAC CTGTAGATAA ATTACCAAAC TTGAAAAAA TTGTTGTCGT 5820  
 30 AGATGAGGAT GTAGAAGACA ATTACATAGA CTTCAATTAGT TTGATGGAAA CTGCTAGCGA 5880  
 TGAATTTGAC ATTGAATGGT TAAAGTCGGA TGATGGTTG ATTTTACATT ATACATCAGG 5940  
 35 TTCTACTGGG CAACCTAAAG GtGTATTGCA TGTTCAACAA GCAATGTTAG TGCACTATAT 6000  
 TTCTGGAAAA TATGTATTAG ATTTACAAGA AGATGATGTT TATTGGTGTA CAGCAGATCC 6060  
 AGGTGCGGTT aCAGGAACAT CTTATGGTAT TTTTGCACCA TGGTTAAATG GCGCTACAAA 6120  
 40 TTGTATAGCT GGTGGTCGCT TTTCGCCAGA ACAGTGGTAT AGTATGATTG AAGATTTTAA 6180  
 AGTGACGATT TGGTATACGG CACCAACAGC TTTAAGAATG TTAATGAGTG CTGGTGACGA 6240  
 TATTGTTGAG AAATATGACT TGTCATCGTT ACGTTGATT CTATCAGTAG GTGAGCCTTT 6300  
 45 AAATCCTGAA GTTATAAAAT GGGCGAAAAA AGTATACGGT TTAACGGTGT TAGATACTTG 6360  
 GTGGATGACA GAAACAGGTG GACATATGAT TGTAACTAT CCAACGATGG ACGTCAAGCT 6420  
 TGGCTCAATG GCGAAACCAT TACCTGGTAT TCAAGCTGCA ATTATCGATG ATGCAGGGAA 6480  
 50 TGAATTACCA CCAAATCGAA TGGGCAACCT TGCTATAAAA AAAGGCTGGC CATCAATGAT 6540  
 GTATCGTATC TGAAGAATC CAGAAAAATA TAAATCATAT TTTATTGGAG ACTGGTATGT 6600

55

	TGATGTAATT ATGACAGCTG GTGAACGAGT TGGACCATT T GAGGTTGAGT CTAAATTGGT	6720
	TGAACACGAA GCAGTTGCCG AAGCAGGAAT TATTGGTAAA CCTGATCCGG TTCGCGGTGA	6780
5	AATAATTAAG GCGTTTGTG CACTGAGAAA AGGATATGAA CCAACAGACG AATTAAAAGA	6840
	AGAAATTCGT ATATTTGTTA AAGAAGGTTT GTCGGCACAT GCAGCACCAC GTGAAATCGA	6900
10	ATTTAAAGAT AAATTACCTA AAACACGGTC AGGTAAAATT ATGAGACGTG TATTAAAAGC	6960
	TTGGGAATTA AATTTAGATG CTGGGGATT T AAGTACAATG GAATAATGAC ATGAATGTTA	7020
	TTGAAGATTT TTTTCGAAGA ATAAAGGTG ACAACATATT TCATGTCAAT GTTTAAATAA	7080
15	TCGTTTACTT TACGATAAGC AATATAAGA ACTGTAACT TGTGTCATAT CATTTCTAG	7140
	AAAGCATTG AAAATGATGA CATAACAATA ATGGCATATC TTTATATTGC TTTTATTTT	7200
	TAATATGATC TTTGGAAGAT GATTATTTTA AATAATAGAA AAATATAGTT ATCAATAGTA	7260
20	TCAAGCGCTA AAAGTTGTAT AATACAAAAC TTTAATAAGT GAATTTATTG CAAAAATGAA	7320
	AGCGCTAACC CGATTTAGTC GACAAGTTTT TAACAGTTCG TTATTATATG AATGTAAGTA	7380
	AAAATTTCTT AGCTACAAC TACATATTAT AAATGCATAA ATTAAACAAA AAGGGCGAA	7440
25	AAAAGTTGAC TCATTTATCA GATTAGATA TTGCGAATCA ATCAACACTA CAACCAATTA	7500
	AGGATATTGC TGCATCAGTA GGTATTTT CAG AGGATGCATT AGAACCTTAT GGTCAATACA	7560
30	AAGCTAAAAT CGACATTAAT AAAATTACGC CAAGAGAAAA CAAAGGGAAA GTTGTTTTAG	7620
	TAAGTGCAT GAGCCCAACA CCAGCTGGTG AAGGTAAATC AACGGTTACA GTTGGTTTAG	7680
	CTGATGCATT CCATGAGTTA AATAAAAACG TTATGGTTGC ATTAAGAGAG CCTGCTTAG	7740
35	GACCAACATT TGGTATCAAA GGTGGTGCGA CTGGTGGTGG TTATGCGCAA GTCTTACCTA	7800
	TGGAAGATAT CAACTTACAT TTCAACGGAG ATTTCCATGC GATTACAAC T GCAAATAATG	7860
	CATGCTCTGC GTTTATCGAT AATCATATTC ACCAAGGTAA CGAATTAGGA ATCGATCAAA	7920
40	GACGTATTGA GTGGAAACGT GTATTAGATA TGAATGATCG TGCACCTAGA CATGTAAACG	7980
	TTGGGTTAGG TGGACCTACA AATGGTGTAC CACGTGAAGA TGGCTTTAAT ATTACAGTAG	8040
	CGTCTGAAAT TATGGCGATT TTATGTTTAA GTAGAAGTAT TAAAGACTTA AAAGATAAAA	8100
45	TTAGTCGTAT TACTATTGGT TACACTAGAG ATCGCAAGCC AGTTACAGTT GCAGATTTAA	8160
	AAGTGGAAGG TGCACCTGCA ATGATTTTAA AAGATGCAAT AAAACCAAAC TTAGTACAAT	8220
50	CAATTGAAGG GACACCTGCA TTAGTTCATG GTGGACCATT TGCGAATATC GCACACGGTT	8280
	GTAAGTCAAT TTTAGCAACT GAAACAGCAC GTGATTTAGC TGATATCGTT GTAACGGAAG	8340
	CTGGATTTGG TTCAGACTTA GGCGCTGAAA AATTCATGGA CATTAAAGCG CGTGAAGCAG	8400
55		

	GTGTAGCGAA AGATAATTTA AAAGAAGAAA ATGTAGAAGC AGTAAAAGCA GGAATTGTTA	8520
	ATTTAGAGCG TCATGTTAAT AATATTAAAA AATTCGGTGT AGAACCGGTT GTTGCAATTA	8580
5	ATGCATTAT ACATGATACC GATGCAGAAG TAGAATATGT AAAATCTTGG GCTAAAGAAA	8640
	ATAACGTACG AATTGCCTTA ACTGAAGTTT GGGAAAAAGG TGGTAAAGGT GGC GTTACT	8700
	TAGCAAATGA AGTATTAGAA GTCATTGATC AACCTAATTC ATTTAAACCT TTATATGAAT	8760
10	TAGAATTACC ATTAGAGCAA AAGATTGAAA AGATTGTGAC TGAAATCTAT GGCGGTTCAA	8820
	AAGTAACGTT TAGCAGTAAA GCGCAAAAAC AATTAATAACA ATTTAAAGAA AATGGTTGGG	8880
15	ATAATTACCC AGTATGTATG GCGAAAACAC AATATTCATT CTCAGATGAT CAAACGTTGT	8940
	TAGGTGCACC ATCAGGATTT GAAATTACAA TTCGTGAATT AGAAGCGAAA ACAGGTGCAG	9000
	GATTTATCGT AGCGTTGACA GGTGCAATCA TGA CTATGCC TGGTTTACCT AAAAAACCAG	9060
20	CAGCATTAAA CATGGATGTT ACTGATGATG GTCATGCAAT TGGGTTATTC TAATAAATCA	9120
	TGTCAATTGT TTAATAAAGA TAAGTAAATA GTTTAATAGA CCGGACTGTT GGAGATGCAT	9180
	TATTTAGCA GTTCGGTTT TTGCTGTGCT AAAAATAGAT TCAATTGGC GAATCTAACG	9240
25	ACAATGTTTG AAGGTGGTTA ATTAATGTAT ATGAAGATAA AAAGTGGGCT TGAAGAATAG	9300
	GAAAGCGATG CAATGAATAT TCCATATTAA AAAAAATTAA TAAAATAGGT TGCAATATTT	9360
	AATTGGGATG CGCTACAATT AACACTAATA ATTGATATTG ATAATTATTA TCAATTAAAT	9420
30	ATAATCTTAT AGGAGTTGTT AACACATGA ACAACATCA CCCAAAATTA AGGTCTTTCT	9480
	ATTCTATTAG AAAATCAACT CTAGGCGTTG CATCGGTCAT TGTCAGTACA CTATTTTAA	9540
35	TTACTTCTCA ACATCAAGCA CAAGCAGCAG AAAATACAAA TACTTCAGAT AAAATCTCGG	9600
	AAAATCAAAA TAATAATGCA ACTACAATC AGCCACCTAA GGATACAAAT CAAACACAAC	9660
	CTGCTACGCA ACCAGCAAAC ACTGCGAAAA ACTATCCTGC AGCGGATGAA TCACTTAAAG	9720
40	ATGCAATTAA AGATCCTGCA TTAGAAAATA AAGAACATGA TATAGGTCCA AGAGAACAAG	9780
	TCAATTTCCA GTTATTAGAT AAAACAATG AAACGCAGTA CTATCACTTT TTCAGCATCA	9840
	AAGATCCAGC AGATGTGTAT TACTATAAAA AGAAAGCAGA AGTTGAATTA GACATCAATA	9900
45	CTGCTTCAAC ATGGAAGAAG TTTGAAGTCT ATGAAAACAA TCAAAAATTG CCAGTGAGAC	9960
	TTGTATCATA TAGTCCTGTA CCAGAAGACC ATGCCTATAT TCGATTCCCA GTTTCAGATG	10020
	GCACACAAGA ATTGAAATTT GTTCTTTCGA CTCAAATTGA TGATGGAGAA GAAACAAATT	10080
50	ATGATTATAC TAAATTAGTA TTTGCTAAAC CTATTTATAA CGATCCTTCA CTTGTAAAT	10140
	CAGATACAAA TGATGCAGTA GTAACGAATG ATCAATCAAG TTCAGTCGCA AGTAATCAAA	10200

55

	AGGCAACGAC CAATATGAGT CAACCTGCAC AACC AAAATC GTCAACGAAT GCAGATCAAG	10320
	CGTCAAGCCA ACCAGCTCAT GAAACAAATT CTAATGGTAA TACTAACGAT AAAACGAATG	10380
5	AGTCAAGTAA TCAGTCGGAT GTTAATCAAC AGTATCCACC AGCAGATGAA TCACTACAAG	10440
	ATGCAATTAA AAACCCGGCT ATCATCGATA AaGAACATAC AGCTGATAAT TGGCGACCAA	10500
	TTGATTTTCA AATGAAAAAT GATAAAGGTG AAAGACAGTT CTATCATTAT GCTAGTACTG	10560
10	TTGAACCAGC AACTGTCATT TTTACAAAAA CAGGACCAAT AATTGAATTA GGTTTAAAGA	10620
	CAGCTTCAAC ATGGAAGAAA TTTGAAGTTT ATGAAGGTGA CAAAAAGTTA CCAGTCGAAT	10680
15	TAGTATCATA TGATTCTGAT AAAGATTATG CCTATATTCG TTTCCCAGTA TCTAATGGTA	10740
	CGAGAGAAGT TAAATTTGTG TCATCTATTG AATATGGTGA GAACATCCAT GAAGACTATG	10800
	ATTATACGCT AATGGTCTTT GCACAGCCTA TTACTAATAA CCCAGACGAC TATGTGGATG	10860
20	AAGAAACATA CAATTTACAA AAATTATTAG CTCCGTATCA CAAAGCTAAA ACGTTAGAAA	10920
	GACAAGTTTA TGAATTAGAA AAATTACAAG AGAAATTGCC AGAAAAATAT AAGGCGGAAT	10980
	ATAAAAAGAA ATTAGATCAA ACTAGAGTAG AGTTAGCTGA TCAAGTTAAA TCAGCAGTGA	11040
25	CGGAATTTGA AAATGTtACA CCTACAAATG ATCAATTAAC AGATTTACAA GAAGCGCATT	11100
	TTGTTGTTTT TGAAAGTGAA GAAATAGTG AGTCAGTTAT GGACGGCTTT GTTGAACATC	11160
	CATTCTATAC AGCAACTTTA AATGGTCAAA AATATGTAGT GATGAAAACA AAGGATGACA	11220
30	GTTACTGGAA AGATTTAATT GTAGAAGTA AACGTGCAC TACTGTTTCT AAAGATCCTA	11280
	AAAATAATTC TAGAACGCTG ATTTTCCCAT ATATACCTGA CAAAGCAGTT TACAATGCGA	11340
35	TTGTTAAAGT CGTTGTGGCA AACATTGGTT ATGAAGGTCA ATATCATGTC AGAATTATAA	11400
	ATCAGGATAT CAATACAAAA GATGATGATA CATCACAAAA TAACACGAGT GAACCGCTAA	11460
	ATGTACAAAC AGGACAAGAA GGTAAGGTTG CTGATACAGA TGTAGCTGAA AATAGCAGCA	11520
40	CTGCAACAAA TCCTAAAGAT GCGTCTGATA AAGCAGATGT GATAGAACCA GAGTCTGACG	11580
	TGGTTAAAGA TGCTGATAAT AATATTGATA AAGATGTGCA ACATGATGTT GATCATTTAT	11640
	CCGATATGTC GGATAATAAT CACTTCGATA AATATGATTT AAAAGAAATG GATACTCAAA	11700
45	TTGCCAAAGA TACTGATAGA AATGTGGATA AAGATGCCGA TAATAGCGTT GGTATGTCAT	11760
	CTAATGTGCGA TACTGATAAA GACTCTAATA AAAATAAAGA CAAAGTCATA CAGCTGAATC	11820
	ATATTGCCGA TAAAAATAAT CATACTGGAA AAGCAGCAAA GCTTGACGTA GTGAAACAAA	11880
50	ATTATAATAA TACAGACAAA GTTACTGACA AAAAAACAAC TGAACATCTG CCGAGTGATA	11940
	TTCATAAAAC TGTAGATAAA ACAGTGAAAA CAAAAGAAAA AGCCGGCACA CCATCGAAAG	12000
55		

CATGGTGGGG CTTATATGCG TTATTAGGTA TGTTAGCTTT ATTCATTCCT AAATTCAGAA 12120  
 AAGAATCTAA ATAATTAnCT AAATATAGCA TATGTATGAT TAACTTTGTA GAC 12173

5

(2) INFORMATION FOR SEQ ID NO: 311:

10

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1316 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

20

25

30

35

40

45

50

55

CAACATTAAT ATTGATATTA AATCTTCCTG GATAACGTGC TTGTTGAGT GATAAGTATG 60  
 CACGCACCTG ACTTAACTCT TTATCTAAAG TAATCGTATG TTGCTTAGAG CCTTGTAAT 120  
 TCGCTCTGAA AAAATAACTC AATTCTAATA GTAACCTCG TGCCTTTTCG CTATTTATTC 180  
 TAACTAAAGC TGAGATCGTG TTAATTGAAT TGAAGAAAAA ATGTGGACTC ACTTGTGCCT 240  
 GTAATGACTT AATCTCAGCA TCTTTCAATA ACTTACTTTG CGTTTCGGCT TCACCAAGTT 300  
 CAATTTGGCT ACTAAAAATA TTTGCCAATC CTTCTGCAAG TTGACGTTCC ACAAAGTTA 360  
 AATCATTAGG GTTTGTAAAA TACATCTTCA ATGTACCGAC GATAGAACCA TGCATCTCAA 420  
 GTGGTATCAC GATAGCTGCT CTAAGCGGGC AATTCGGATG ACTACAACCA ATCTCTTCTT 480  
 TAGTATGAAC TTCTTTCAAC TTTCCTGATT TCAATACATC TTTAGACAGA CTTGTTAATA 540  
 TTTCATTTGT TGGTATGTGA TGATCACTAC CTGCACCTAC ATGCGATAAG ATTTCATTTT 600  
 TGCTTGTAAT TGCTACGGCA GATACTTTCA TTAAATTTTT AATAATCATC GCAATTTGCT 660  
 GTGCCGATTC TCTATTCAAT CCTTCTTTAA AATACGGCAA TGTCTGGTTC ATCAATTGCA 720  
 GTAATCATG TGTTTGAACA GCCTTCATTT GCTCCTCTTG CTTTAATGTT GAAATGATAA 780  
 TAGACATAAA AATCGCCGTA CCAACGCTAT TAACAATAAT CATTGGTAGT GCAATTAATG 840  
 ATATGAGGTC AACCGCATAT GCTTTGTCGT GGGAAAATGT TAAAATGCTC AACATTTGAA 900  
 TCATTTCCAT AACAATTCCA ATCATGGCAC TTTTCGCAAT ACTCGGTAA CGCTTGCCTC 960  
 TTTGAGCTTG TAAGCCAAAA TAACCAGCAA TTATACCAAT AAATATAGAT GAGATAAGAT 1020  
 AAACCTGTGC ATCCGCCCA CCCATATACA CTCTGAAAAT ACCTGAAATA ACGCCAACAA 1080  
 ATAGACCTAC AAAAGGGCCA CCAACTAATC CTGCGACACC TATCGTTAAT ACACGTGTGT 1140  
 TAGCTAAAGA TACATCATCA TCTAAACGGA AGTACACACT TCCTGACAAA CTATGTTGAT 1200  
 GATCGATGAC GATACCAGTT AAATTAGACA TTAAGGCAAA CAACTGAAA ATAATACATA 1260

## (2) INFORMATION FOR SEQ ID NO: 312:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7972 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

TATAAATATT ATTTTATTAT CGTTTATCGC TATTTGTGTT ACATAATCAA AACCATAAAT 60  
 TCTTACTCAT TCAGATTTAC CCAATATTTT TACTTTTATA ATGTAATGCG TTTTATCCAA 120  
 GTTATTTTTT AAAAATAAAT ATTGAATTnG GGGCTGnTTT CATGTCATTA AGAGATGAAG 180  
 CATTGGAAT GCACAAACGT AATCAAGGTA AATTAGAAGT TAAACCAAAT GTAAAAGTTA 240  
 CTAATAAAGA GGAATTAAGT TTAGCATACT CACCTGGCGT TGCTGAACCG TGTAAAGATA 300  
 TTTATGAAGA TAAAAGAAAA GSTATGATT ACACAATTAA AGGAAATACA GTTGCAGTTA 360  
 TTACTGATGG AACAGCGGTA TTAGGTTTAG GTAACATTGG ACCTGAAGCA AGTATTCCTG 420  
 TAATGGAAGG TAAAGCAGTA TTATTCAAAA GCTTCGCTGG TATCAATGGG GTGCCTATTG 480  
 CGTTAAATAC AACTGATACC GAAGAAaTCA TTAAACAGT TAAGTTGTTA GAACCTAATT 540  
 ATGGTGGTAT TAATTTAGAG GATATTTTCGG CACCACGTtG TTTTGAAATT GAAGAACGAT 600  
 TGAAAAAAGA AACTAATATT CCGGTATTCC ATGACGATCA ACATGGTACA GCAATTGTAA 660  
 CATTGGCAGG TTTGGTAAAT GCATTGAGAG TTGTTAACAA AGATATTGCT AAAATAAAAG 720  
 TTGTACTAAA TGGTGCTGGT GCAGCAGGAA TAGCCATTGT TAAATTACTA TACGCGTATG 780  
 GTGTAAGAAA TATGGTTATG TGTGACTCAA GAGGCGCAAT TTTTGAAGGA CGTTCATATG 840  
 GTATGAATCC TACGAAAGAT GTTGTAGCAA AATGGACAAA TAAAGATAAG ATTGAAGGGT 900  
 CTTTAGAAGA AGTCGTA AAA GACGCAGATG TATTTATCGG GGTTCCTGTA GCTAATGCGC 960  
 TGTCACAAGA TATGGTTAAG AGTATGGCAG ATAATCCAAT TATATTTGCA ATGGCTAATC 1020  
 CAAATCCTGA AATAATACCT GATGATGCCA AAGCGGCAGG TGCACGAGTT GTTGGTACAG 1080  
 GACGTTTACA CTATCCTAAC CAAATTAATA ATGTATTAGC TTTCCCTGGT ATTTTATAGAG 1140  
 GTGCATTAGA GGTGAAGCT ACACATATAA ATGAAGAAAT GAAAAAGGCA GCTGTAGAAG 1200  
 CGATTGCTGA TTTAATCGAT AGTTCTGAAT TAAATGAAGA CTACTGTATC CCAGGACCGT 1260  
 TTGATAAACG TGTAGCGCCA TCAGTTGCTC GTAATGTTGC TAAAGCGGCA ATGGAATCTG 1320  
 GAGTAGCTAG GATTGAAGTT GATCCGCAAG ATGTGTATGA TAAAACAATG AAACCTACAG 1380

	ATTAAATGA TGAATGAA AGTTTATGAT AAACATTCAA CAGTCAAACG AATATAAATC	1500
	AAATAAATTT AAACCCGTTT TTAAGTGGTC AAGTTCAGTT TAAGGCTCTA AATGGTTAGA	1560
5	ACAGAGGTTA TTTGGAGGTT TTCCTATGTT TAAAGATTTT TTTAATCGAA CAAAGAAAAA	1620
	GAAATATCTT ACAGTACAAG ACTCTAAAAA TAATGATGTG CCTGCAGGTA TTATGACTAA	1680
10	GTGTCCAAAG TGTAAGAAAA TTATGTACAC AAAAGAATTA GCTGAAAATT TAAATGTGTG	1740
	CTTTAATTGT GATCATCATA TTGCTTTAAC TGGCTATAAA CGTATAGAAG CAATTTCTGA	1800
	TGAAGGATCA TTTACAGAAT TCGATAAGGG AATGACCTCT GCGAATCCAT TAGATTTTCC	1860
15	aAGTTATTTA GAAAAAATTG AAAAGGACCA ACAAAGACA GGTCTTAAAG AAGCAGTTGT	1920
	GACTGGTACA GCACAAC TAG ATGGTATGAA ATTTGGCGTT GCTGTCATGG ATTCACGTTT	1980
	TAGAATGGGA AGTATGGGAT CGGTTATCGG TGAAGAGATA TGTGCGATCA TTGATTACTG	2040
20	CACTGAGAAC CGTTTACCAT TTATTCTTTT CTCTGCAAGT GGTGGTGCAC GTATGCAAGA	2100
	AGGTATTATT TCCTTGATGC AAATGGGTAA AACCAGTGTA TCTTTAAAC GTCATTCTGA	2160
	CGCTGGACTA TTATATATAT CATATTTAAC ACATCCAAC ACTGGTGGTG TATCTGCAAG	2220
25	TTTTGCATCA GTTGGTGATA TAAATTTAAG TGAGCCAAAA GCGTTGATAG GTTTTGCAGG	2280
	TCGTCGAGTT ATTGAACAGA CAATAAACGA AAAATTGCCA GATGATTTC AACTGCAGA	2340
30	ATTTTTATTA GAGCATGGAC AATTGGATAA AGTTGTACAT CGTAATGATA TGGCTCAAAC	2400
	ATTGTCTGAA ATTCTAAAAA TCCATCAAGA GGTGACTAAA TAATGTTAGA TTTTGAAAAA	2460
	CCACTTTTTG AAATTCGAAA TAAAATTGAA TCTTTAAAG AATCTCAAGA TAAAAATGAT	2520
35	GTGGATTTAC AAGAAGAAAT TGACATGCTT GAAGCGTcAT TGGAACGAGA AACTAAAAAA	2580
	ATATATACAA ATCTAAAACC ATGGGATCGT GTGCAAATTG CGCGTTTGCA AGAAAGACCT	2640
	ACGACCCTAG ATTaTATTCC ATATATCTTT GATTCGTTTA TGGAACTACA TGGTGATCGT	2700
40	AATTTTAGAG ATGATCCAGC AATGATTGGT GGTATTGGCT TTTTAAATGG TCGTGCTGTT	2760
	ACAGTTaTTG GACAACAACG TGGAAAAGAT ACAAAGATA ATATTATCG AAATTTTGGT	2820
	ATGGCGCATC CAGAAGGTTA TCGAAAAGCA TTACGTTTAA TGAACAAGC TGAAAAATTC	2880
45	AATCGTCCTA TCTTTACATT TATAGATACA AAAGTGCAT ATCCTGGTAA AGCTGCTGAA	2940
	GAACGTGGAC AAAGTGAATC TATCGCAACA AATTTGATTG AGATGGCTTC ATTAAAAGTA	3000
50	CCAGTTATTG CGATTGTCAT TGGTGAAGGT GGCAGTGGAG GTGCTCTAGG TATTGGTATT	3060
	GCCAATAAAG TATTGATGTT AGAGAATAGT ACTTACTCTG TTATATCTCC TGAAGGTGCA	3120
55	GCGGCATTAT TATGGAAAGA CAGTAATTTG GCTAAAATTG CAGCTGAAAC AATGAAAATT	3180



	GGTGCACATA AAGATATTGA ACAGCAAGCT TTAGCTATTA AATCAGCGTT TGTTCACAG	3300
	TTAGATTCAC TTGAGTCATT ATCACGTGAT GAAATTGCTA ATGATCGCTT TGAAAAATTC	3360
5	AGAAATATCG GTTCTTATAT AGAATAATCA ACTTGAGCAT TTTTATGTTA AATCGATACT	3420
	GGGTTTACC ATAAATTGAA GTACATTAAA ACAATAATTT AATATTTAGA TACTGAATTT	3480
	TAAC TAAGAT TAGTAGTCAA AATTGTGGCT ACTAATCTTT TTTTAATTAA GTTAAAATAA	3540
10	AATTCAATAT TTAAACGTT TACATCAATT CAATACATTA GTTTTGATGG AATGACATAT	3600
	CAATTTGTGG TAATTTAGAG TTAAAGATAA ATCAGTTATA GAAAGGTATG TCGTCATGAA	3660
15	GAAAATTGCA GTTTTAACTA GTGGTGGAGA TTCACCTGGA ATGAATGCTG CCGTAAGAGC	3720
	AGTTGTTCGT ACAGCAATTT ACAATGAAAT TGAAGTTTAT GGTGTGTATC ATGGTTACCA	3780
	AGGATTGTTA AATGATGATA TTCATAAACT TGAATTAGGA TCAGTTGGGG ATACGATTCA	3840
20	GCGTGGAGGT ACATTCTTGT ATTCAGCAAG ATGTCCAGAG TTTAAGGAGC AAGAAGTACG	3900
	TAAAGTTGCA ATCGAAAAC TACGTAAAAG AGGGATTGAG GGCCTTGTAG TTATTGGTGG	3960
	TGACGGTAGT TATCGCGGTG CACAACGCAT CAGTGAGGAA TGTAAGAAA TTCAAACATAT	4020
25	CGGTATTCCT GGTACGATTG ACAATGATAT CAATGGTACT GATTTTACAA TTGGATTGTA	4080
	CACAGCATTA AATACGATTA TTGGCTTAGT CGACAAAATT AGAGATACTG CGTCAAGTCA	4140
	CGCACGAACA TTTATCATTG AAGCAATGGG CCGTGATTGT GGAGATCTAG CATTATGGGC	4200
30	TGGATTATCA GTTGGTGCTG AGACAATTGT AGTTCCAGAA GTGAAAACAG ATATTAAAGA	4260
	AATAGCTGAT AAAATTGAAC AAGGTATTAA ACGTGGTAAG AAACACTCAA TCGTTCTTGT	4320
35	AGCAGAAGGT TGTATGACTG CGCAAGATTG TCAAAAAGAA TTATCACAAT ACATCAATGT	4380
	TGATAATAGA GTGTCTGTGT TAGGTCACGT TCAACGTGGT GGTAGCCCAA CAGGTGCGGA	4440
	TAGAGTTTTA GCATCACGTT TAGGTGGATA TGCGGTAGAC TTATTAATGC aAGGTGAAAC	4500
40	AGCTAAGGGT GTTGGAATTA AGAACAATAA AATTGTAGCA ACATCTTTTG ATGAAATTTT	4560
	TGATGGTAAA GATCATAAAT TTGATTATAG TCTATATGAA CTGCTAACA AGTTATCTAT	4620
	ATAAGATTTC AGGAGGAATT ATAAAATGAG AAAAATAAA ATTGTATGTA CAATTGGACC	4680
45	AGCTTCAGAA TCAGAAGAAA TGATTGAGAA ATTAATCAAT GCTGGTATGA ACGTTGCACG	4740
	ATTAACTTT TCACATGGTA GTCATGAAGA GCATAAAGGT AGAATTGATA CAATTGTA	4800
	AGTAGCTAAA AGATTAGACA AAATTGtAGC AATTTTATTA GATACAAAAG GTCCAGAAAT	4860
50	TCGTACGCAT AATATGAAAG ACGGTATCAT TGAAGTTGAA CGTGGCAACG AAGTTATTGT	4920
	TAGCATGAAT GAAGTTGAAG GAACACCTGA AAAGTTCTCA GTAACATATG aAAACTTAAT	4980
55		

	TAAAGATATT GACCATGCTA AAAAAGAAGT TAAATGTGAT ATTTTAAACT CTGGTGAGCT	5100
	TAAAAACAAA AAAGGTGTTA ACTTACCTGG CGTAAGAGTA AGTTTACCTG GTATTACAGA	5160
5	AAAAGATGCT GAAGATATCC GTTTCGGTAT TAAAGAAAAT GTTGACTTCA TTGCAGCAAG	5220
	TTTCGTACGT CGTCCTAGTG ATGTTTTAGA AATTCGTGAA ATTTTAGAAG AACAAAAAGC	5280
	TAACATTTCA GTATTCCCTA AAATTGAAAA CCAAGAAGGT ATTGATAATA TTGCGGAAAT	5340
10	TCTTGAAGTG TCTGATGGTT TAATGGTTGC ACGTGGTGAC ATGGGTGTTG AAATTCCACC	5400
	TGAAAAAGTA CCAATGGTTC AAAAAGATTT AATCAGACAA TGTAACAAAT TAGGTAAACC	5460
15	AGTTATTACA GCTACACAAA TGTTAGATT C TATGCAACGT AACCCACGTG CTACACGTGC	5520
	AGAAGCTAGT GACGTTGCCA ACGCAATCTA TGATGGTACA GATGCAGTAA TGTTATCTGG	5580
	TGAAACTGCT GCTGGTTTAT ATCCTGAAGA AGCTGTTAAA ACAATGAGAA ATATTGCTGT	5640
20	ATCAGCTGAA GCAGCCCAAG ATTACAAAAA GTTATTGTCA GATCGTACTA AATTAGTTGA	5700
	AACTTCATTA GTGAATGCTA TCGGTATTTT GGTTCACAT ACAGCTTTAA ACTTAAATGT	5760
	TAAAGCAATT GTAGCTGCTA CTGAAAGTGG TTCAACGGCA CGTACTATCT CCAAATATCG	5820
25	TCCACATTCA GACATTATTG CGGTGACTCC AAGTGAAGAA ACTGCACGTC AATGTTCAAT	5880
	TGTTTGGGGA GTTCAACCTG TAGTTAAAAA AGGACGTAAG AGTACAGATG CATTGTTAAA	5940
	CAATGCAGTT GCAACAGCTG TTGAAACTGG TAGAGTATCT AATGGTGATT TAATCATTAT	6000
30	TACTGCTGGT GTACCAACTG GTGAAACTGG AACTACTAAT ATGATGAAAA TCCACCTAGT	6060
	TGGTGACGAA ATTGCTAATG GTCAAGGTAT TGGACGTGGA TCAGTTGTTG GTACTACGTT	6120
35	AGTTGCTGAA ACTGTTAAAG ATTTAGAAGG TAAAGATTTA TCTGACAAAG TTATCGTTAC	6180
	TAACCTCAATC GATGAAACGT TTGTACCTTA TGTAGAAAAA GCTTTAGGCT TAATTACAGA	6240
	AGAAAATGGT ATTACATCAC CAAGTGCAAT TGTTGGTTTA GAAAAAGGTA TTCCAACAGT	6300
40	TGTAGGTGTA GAAAAAGCTG TTAAAAACAT AAGCAATAAC ATGTTAGTTA CGATTGATGC	6360
	TGCTCAAGGT AAAATCTTTG AAGGATATGC AAACGTACTA TAATTTATAA AAAAACGTCT	6420
	TTCCATTTAT CAACAATGGA AAGGCGTTTT TTGGTTcATC TGGTATTTTA TGACGTAATT	6480
45	AATAGGTTAT TTGATAATGA TAGTGTATGA ATGGCAATCT ATATAAATGT TTATATCTTT	6540
	TATACATGTA CATTATCACC TTCAAACCTT CACTCATATT ACTTTGGAAA TTTATTATAA	6600
	AATAGAAGTA TGGATGTATT TCTGAAATGA TACATTATTA AATAGATGAG AAAGTAAAAG	6660
50	TTTTGAGCCA AGTACGCAAT TTAATATTAT AAGTTGCATA TAAAACAGGA TGGGACATAA	6720
	ATCCCTAAAA AAACAGCAGT AAGATAATTT TCAATTAGAA AATATCTTAC TGCTGTTCTC	6780
55		

5      tCTTCGACTG GCACTGCTCC CTCAGGAGTC TCGCCATTAA TACTACGTAT TAACATGTAA 6900  
 TTTTACTTTT ACATACTTTA AAAAAATAAA ACACTTTGCC CAACTTGCAC ATAAATGTAA 6960  
 AATTCAATAA AATGAATTTT CTGTGTTGGG TCCCTTCGTA TAATTTAATA AATACCACTA 7020  
 AACTAAATTA ACGAGGTGCC TTATGTATAA AATTTATAAC ATGACCCAAC TTACTACTACC 7080  
 10      AATAGAAACC TCTGTTAGAA TTCCTCAAAA TGATATTTCC CGATATGTTA ATGAAATTGT 7140  
 TGAAACGATA CCTGATAGCG AATTCGATGA ATTCAGACAT CATCGTGGCG CAACATCCTA 7200  
 TCATCCAAAA ATGATGTAA AAATCATCTT ATATGCATAT ACTCAATCTG TATTTTCTGG 7260  
 15      TCGTAGAATA GAATAATTAC TTCATGACAG TATTCGAATG ATGTGGTTAG CTCAAGATCA 7320  
 AACACCTTCT TATAAACTA TTAATCGTTT TAGAGTGAAT CCTAATACTG ATGCGTTAAT 7380  
 TGAATCTTTA TTTATTCAGT TCCATAGTCA ATGTTTAAAG CAAAATCTTA TTGATAATAA 7440  
 20      TTCAATTTTT ATTGATGGTA CAAAAGTAGA AGCTAATGCC AATAGATATA CATTTGTGTG 7500  
 GAAGAAAAGT ATTCAAAATC ACGAATCGAA ATTGAACGAA AATCAAAAA CATTATATCG 7560  
 TGACTTAGTT GAAGAAAAAA TAATACCAGA GATAAAAGAA GATGGAGATA GCGATTTAAC 7620  
 25      AATAGAAGAA ATAGATTTAA TTGGTAGTCA TTTAGATAAA GAAATCGAAG ATTTAAATCA 7680  
 TTCTATTGAG AACGAAGATT GTGCTCAAAT TAGAAAACAG ACCCGTAAAA AAATAACTGA 7740  
 GATTAAGAAG TTCAAAAAGA AATTGATGA TTATTCCGAA AGAAAAATA AATATGAAGA 7800  
 30      ACAAAAATCG ATTCTTAAAG ATAGAAATAG TTTTCTAAA ACTGATCTGA TCATGATGCA 7860  
 ACTTTTATGA GAATGAAGGA AGACCATATG AAAAATGGCC AACTTAAGCC AGGATACAAT 7920  
 35      TTACAAATAG CGACAAATTC TCAAAAATGT TTTATCCTAT GACCTATTTT AA 7972

## (2) INFORMATION FOR SEQ ID NO: 313:

40      (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3175 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

50      ATTTTTTAGT TAATTGTCTT TCTTAAATA ATTTTAGCTT TCATTAAATT AAACAATTTT 60  
 ACAAGCTTGG AACACCAATC AAAATCCTAA GTTCTAAAT GCAATATTAG TAGTCGTGA 120  
 CTGAATGAAC ATATGCTTAT AATATTTTTT TGCAATGCTA GTCAAGTTGA TTTATGCTCA 180  
 CAAGGATATG CGATTTATAT TTTCTTACAA CAATGAAAAT GCCTGATACA ATGCGATCCT 240

55

	AATTAAATTA TTGTTGTTTT AACCATTGTA CTATCGTCGG AATCATAAAT CCTGTTGGCC	360
	CTTTTGGACC ATTATATGAA GCTTTATTAG TCGTTGCTGG ACCTGCAATA TCGAAATGAA	420
5	TGTGAGGTGT TTGACCACTA AAATGTGTTA CAAAACCTGC CGCAAATAGC GCTTTACCTT	480
	GTCCATTTCGT ATGGTTAACT AAATCAGCGA TATCACTGTG TTTAATACTT GCACGTTCGG	540
	TTGCAGTAAT CGGTAATTCA AATACCATTT CATCGACTTC AGAACTTATT TGTAATATAT	600
10	CGTTTAATAT CACTTTACTA TTCGATTCAA ATGCAGCAGC TTTATCATCG CCTAGTGCAa	660
	CAATTGCTGC ACCCGTTAAT GTAGCAAAGT CCATAATCAC ACTAGGCTGA TATTGATTTG	720
15	CATAAAACAC AGCATCTGCA AGGACTAATC TACCTTCAGC GTCTGTATTC ATTACTTCTA	780
	CAGTTTCACC ACTTAATGCT GTAAATACAT CATCTGGcTT CATTGATGCT TCATTTATCA	840
	TATTTTCAGC ACACGCAAGC ACTCCGACAA TATTTACAGG CAGTTGTAAA CGAcTAGCCG	900
20	CTTCAATGAT ACCAACGACA TTCGCAGCGC CACACATGTC AAACCTTCATT GTAGCCATGC	960
	CATTCTTCGT TTTAATACTA TAACCACCAG AATCATACGT TATACCTTTA CCAACTAAGG	1020
	CAATTGGTGC TTCATCTTTG TCTTTGCCAT TATATGTGAT GGTTACTAAc CTCGGTTTAT	1080
25	GCTTACTACC TTTACCTACT GCTTGTAATA GTCCGAATCC TTCAGAACT AAAGTGTCA	1140
	AATCTTTAAC ATCTACTTTG ACCTTTGTAT TTTTAAATG ATTAACAATA TCTTCTGcAA	1200
	ATGTTTGTGG TGtTAATACA TTCGGTGGcA TATTACTAAA GTCTCyTGcC AAATTAATGG	1260
30	aTTGGcCAAT ACTGATACCC TCATGAATAA AATCTAATTC AATTAATGAT TCACTAATAA	1320
	GGkTTAAATT CGTCTTAAAC GGtGCCTTCT kACTTgrTTT ATAATGATCG aATyCATATG	1380
35	kTGcACGCTC ACTTTGGAAT ACCGCATGCC ATTAATACAT CTGgATAACT GgATCcATAT	1440
	TTTGgAnATA AATGaATCCa TAAGTAAATA CGTATCTTCa ATGTGTTCTG ACTTTATGTA	1500
	TTGGAAGA TGTCCCCaTA TTTTCAACAT ATCTTGATAA cACGTGTCTT TAAGTTTCCT	1560
40	AAACCAACTG TAATTAAACG ATATGTTTGA TCTTGACAT CAAATGCAGT TGTATAAATC	1620
	TTCCCAACTT TACTACCAAT AATATGTTGA TGTTTTAGTC TTTCAAGTGA TTCTGTAATA	1680
	TCGATATGAT TAAAACTAAT GCGCTCTAAC TGATTTAAAT GTTCTGGTAT ACCAATAATC	1740
45	AATGTATTTA TTTCGTTGCT TAGTGTGTTA TTTAATTTAA AATTCATAAT GTACCTCCTT	1800
	AGATTTGATA TGTTACTCAC TTCAAATTGT ACAACAATAA AGCCCTcAGT GACACTGAAG	1860
	GACTTTATAA ATGAATAATT TAATTTTATG TGTTTAGCTG AAACCTTATT TAAAGCCTAG	1920
50	AATTTACCTT TTTTGAATGC TAAACCGATA CCACCGATTT TGAATACCGC ACGTGATCA	1980
	ATCACTTTTT TCATGAATGC TGCTTTTTTA CCAGCGATAG GTTTACCAAA TACCATACCT	2040
55		

GTTGATTCAC CGTTTAAGAT GCGTTTAATG TTTTTCAGCAA CACTTTCACC TTGTTGCATT 2160  
 GCAATTTGTG CTGTAGTTGG TAATGGACGT TCTTCTCCAG CTGGGATAAA CGCTGAACAG 2220  
 5 TCACCAATAA CAAAAATGTT GTCCTAACCA TTGATTGTTA AATCTTGCTT TGTAACGATA 2280  
 CGTCCACGTT TAACGCCTTC AAATGATTCT TCCATTAATT TACTACCACG TACACCAGCT 2340  
 GCCCATACTG AAGTACCTGC ATTAAATTGT TGTPTTTCAC CATCTACTTC AACTACAAAA 2400  
 10 CCTTTTTTCGT TACAAGCAAC GATTGGTGTA GCAATTTTAA ATTCAACACC GCGGTCTTCT 2460  
 AAGTAGCTAA CTGCGTGGTT AACTAATTCT TCTGAGAACA TTGGTAACAT TTTAGGTGCT 2520  
 GCTTCAACAC AAGTGATTTT AACTTTATTT TGATCCACAC CATATTTGCT ACATAATTCA 2580  
 15 GGAATCTGT CTGTTAATTC ACCTAAGAAT TCAACACCAG TGAATCCAGC ACCACCAACT 2640  
 AAGATAGATA AATCGTTATC ATCTTTTCT TTTGATGCTG CATAGTTAGC AAATTTGTCT 2700  
 20 TCGATATGAC GTGATAATTC ACGTGCTGTG ATAACATTTT CAATTTGGAA AGCATGATCT 2760  
 TTCATACCTT CGATGCCGAA TGTTCACCTA ACGAAACCTA ATGCTACTAC TAAAATATCA 2820  
 AAGTCATAAA TACCTTGATT TGTTCCTACC TTTTTCAGCAT CACGGTCAAT TTTTGTTACT 2880  
 25 TCTGCTGAA CAAAGTTCAC TTTGTCTTTC TTCAAGACAC TTTCCACAGG ATATAATACA 2940  
 TCTTCATAGT TTAGTGTACC TGCTGATGCT TCATGTAACC ATGTGCTTC ATAGTGATAT 3000  
 TCATTTTAT TAATAAGCGT AATTTCTGCT TCTTCTGTTG ATATCGCTTT TTGCAATTTA 3060  
 30 GTTACAGTTT GTAAACCTGC ATAACCAGCA CCAAGTACAA GTACTTTTTT ACGATCTTGA 3120  
 GCCATTTAAT TnCACCTAAG CnTTCATATT TTTTAAACCA AATGCTGATA ATTAC 3175

## (2) INFORMATION FOR SEQ ID NO: 314:

35

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 702 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

45 CATCATTATT AAAGATTTTC AATCAATACA gAATCACAAT ACGTACGCAT TGTGCACGAT 60  
 AAAAATACAG ATGTGTATAT TAACTATGAA CTACAAGAGC AACTAACGAA CAAAGCTTAC 120  
 ATTGGTGATC ATATTTATGT TGAAGGGATA TGGCTCGAAG TACAAGCTGA TGGTTTAAAT 180  
 50 GTATTGAGTC AGAATACAGT GGCATCGTCA TTAATTCGCT TAACACAAGA GATGCCACAT 240  
 GCACAGGCAG ATGATTACAA TACGTACCAT CGTTCGCCAA GGATTATTCA CCGTGAACCG 300

55

TGGCGTTCCA TTATACCGCC ATTAGTAATG ATTGCTTTAA CTGTTGTCAT CTTTTTAGTG 420  
 AGACCAATTG GTATTTATAT TTTAATGATG ATTGGTATGA GTACAGTAAC GATAGTATTT 480  
 5 GGTATTACAA CGTATTTCTC TGAAGAGAAA AAGTATAACA AAGATGTTGA AAAACGAGAG 540  
 AAAGATTACA AAGCTTATTT GGATAATAAA TCTAAAGAAA TTAATAAAGC GATTAAAGCA 600  
 CAACGTTTTA GTTTGAATTA CCATTATCCA ACGGtTGCTG AAATTAAAGA TATCGTTGAA 660  
 10 ACGAAAGCAC CAAGAATATA TGAaAAAACC ATCGGCATCA TC 702

(2) INFORMATION FOR SEQ ID NO: 315:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 4121 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

TATGTTCCGA CAACGAAACA AAGTGTAATT ACAAGAGCAA AGATAACTTT GAATGTTTGT 60  
 25 AAACGTCCAT CTTTACCTTC AGTTAAATGC ATGAACATTA ATAATTGAxG TCCTGCTTGG 120  
 ACGAATGCAA AGCCAAAGAT AATTGTCAAC TTCGCGTGGA ATGTTAATGA CGTGTATAGT 180  
 GTTACGTAAA CTGCTAAAAG CGTTAATACG ATAGATGCGA TAAATCCTAC AGTATGTTTC 240  
 30 ATTATTGTAC TCATCCGCTA TACACCATCC CTATCATATA TACGGCAGTA AAGATGAAAA 300  
 CCCAAACAAC ATCTAAGAAG TGCCAGTATA AACTTACTAT AAATAATTTT GGCGCATTAT 360  
 ATTTGTCTAA TCCGCGTCGT TGGATTTGGA TTAATAAACA AATGGCCCAA ACGATACCTA 420  
 35 GCGATACGTG ACAACCATGC GTTCCTAATA GGATAAAGAA ACTAGACCAG TAAGAACCAA 480  
 TTGTTGGGTT AACGCCTTCT GATGCATAGT GTGCGAATTC ATAAATTTTCG AATCCAACAA 540  
 40 AGACTAAACC TAAAAGTAAC GTAATGATCA TCCAAAACAT CATTAACTTT TGTTTTTCTT 600  
 GGCGCATGTA GTAAATAGCA ATACCACATG TGTAAGAACT GAATAATAAT GCAAACGTCA 660  
 TTATTAAAAC AAGAGGCAAT TCAAATAACT CAGTAGTCAT TTTACCTGCA TAATCGCCAC 720  
 45 CATGTTGCAA AGTTAATAGT GTCGCAAATA GGGTACCGAA TAACGCAAAT TCGGCTGTAA 780  
 TGAAAATCCA AAAGCCAAGC TTATTTAATT CGCCTTCATG TGTGCGTGAA TCAATAGTGT 840  
 TTGTATCATG ACTCATGACT TACAGCCTCC CTTTCTTTAA TACGCGCTTC TCTTAATCTA 900  
 50 GCTTCAGTTT CTGCAACTTC AGCAGCAGGG ATATGATATC CGTGATCGAT TTGGAAACTG 960  
 CGATAAATCA TAGTACCAAA AATACCGAAT AAACAAATTA ATGCTGGAAT TACAGTTTCG 1020

55

	TTGTTTGGCA TATGAATGTC TTTGTAATTA TGGTTGTCTA AGTAATGACG ACCATGTTCT	1140
	TTCATATCAA CAAATGTGTC GTAGTCATTC CAATCTGGTG TAATGGCAAA GTTGTATTTA	1200
5	GGTGAATTG CTGATGCAGT AGTCCACTCT AGAGTACGAC CAAGGCCATC CCAGTTATCT	1260
	CCAGTTGCTT CACGTGGAGA TTTGAAGTGA CTGTATACGA TACTAACAAC AAGGAATAAG	1320
10	AATCCGATTG CCATTAATAA TGCACCGATA GTTGAAATTA AGTTTAATAA GAACCAACCA	1380
	TCTGATGGCA TATAAGTGTA TAAACGACGT GGCATACCAT CTAATCCAAG AATGAATTGT	1440
	GGTAAGAAAC AAACGTTAAA TCCGATCATG AAGAACCAGA AGCACCATT TTTAATGTT	1500
15	TCGTTTAATT TGTAACCCAT CATCTTTGGA TACCAGAAGA TTAAACCAGC TAAGCAGGCA	1560
	AATACAACAC CAGTAACCAA TGTATAGTGG AAGTGAGCTA CTAAGAAGTA CGTATTGTGA	1620
	TATTGATAGT CAGCTGATGC CATTGCTAAC ATTACACCCG TAACACCACC TAAAAGGAAG	1680
20	TTAGGGATAA ATGCTAATGA GAATAGCATT GGTGACTCAA ATGTAATACG TCCTTTATAT	1740
	AATGTTAATA ACCAGTTAAA CAATTTTACA CCAGTTGGAA TACCAATCAG CATTGTTGAA	1800
	ATTGAGAAGA ATGAGTTGAT TAACGCACCA TTACCCATTG TGAAGAAATG GTGAACCCAA	1860
25	ACTAAGAAAC TAAGGAACGC GATACCGGCA GTTGCCCATTA CCATACTTTG ATGTCCGAAT	1920
	AAACGCTTAC GAGCGAATGT CGGGATAATT TCTGAGTAAA TACCAAATGC TGGAAGGATA	1980
	ACGATATAAA CTTCAGGGTG CCCCCATACC CAGAAGAAGT TAGCCCAAAG CATTGGCATA	2040
30	CCGCCATGTG CAACTGTGAA GAATGCTGTG TCAAATATTC TATCAGTTGT CATTAAATGCT	2100
	AACGCTACTG TTAAAGGAGG GAAAGCAAGA ATAACAATTA ATGTAGTAAT AAATGTTGTT	2160
35	ACTGTAAACA TTGGCATTG CATAAACTTC ATAGTTGGTG TTTTACATCT TAAAATTGTT	2220
	ACAAAGAAGT TGATACCTGT AGCTAAGGTA CCAAGCCCTG AAATTTGTAT AGCTATTAAG	2280
	TAATAGTTAA CACCCGGACC AGGACTGAAT TCACCTGCTA GTGGCGCATA GTTTGTCCAA	2340
40	CCAGCTGCTG GTGAACCACC AATAATAAAT GACAGGTTGA ATAAAATCAT ACCTGCAAAG	2400
	AATAGCCAGA AACTTACGTT GTTTAATACT GGGAATGCAA CATCACGTGC TCCAATTTGT	2460
	AATGGAACAA CGATATTCCA TAAACCAAAG ATAAATGGCA TTGCCATGAA GATAATCATG	2520
45	ATTACACCAT GTGTACTAAA AATTTCTGTTA TAGTGGTTAG ATTCTAAAAA TTTGTTATCA	2580
	GGTACTGTTA ATTGCGCAGC AATAAGTAAC GCATCAATAC CACCACGGAC GAACATTAAT	2640
	ACGGCACAGA TTAAATACAT AATACCGATT TTCTTATGGT CTACAGATGT GAACCATTCT	2700
50	TTGTAAAGAT ATTTCCATAA TTTAAAGTAA GTAATTACTG CGATTAAACC AATAACTAAG	2760
	AATGGGGCAC CAATTTGTGC CATTGTAATC ATcCAGTTAC CTTTAACTAG TAATTGATCC	2820
55		

EP 0 786 519 A2

	TTGAAATTTT CTTCATTCTT TTCGCATTTT TCGATTTCATC TTTCTTGAAC TCATTGTTAT	2940
	ATGGTTCGTC ATTTCCAAGA ATCATCAACT TCATACCATG TCGTTTATAG TTCGCATTTG	3000
5	TAATTTGAGC TTTACGAGCA GGTATTAATG GTTGTCTGA TACATCTTA AACATATTTT	3060
	CTTCACTAGT GAAGTTTGGA TCTTTCAATT CGAAATTGAA ACGTTTATAT GCATAGAAGA	3120
	TGTATTCTGG ATCGGCTGCT GGATCAACAA ACGCCATATG TGTACCATTA AATTCTAAAG	3180
10	CTTTATTAGG TGTGCTTGGT AATAATTGTT TATCAAATGT ATCTTGATCT AACGTTTTCT	3240
	TACCTTTAAC TTCTTTCACC CATTTGTCGT AGTCTTTTTG ACTAACGGCA TTTACTTTAA	3300
	ATGTTTGACG TGAGAATCCT TCACCATTGA AGTTAGAGTT ACGACCTCTG AACGTACCAG	3360
15	TTTGAGATGC TTCTAACGTC CAATTCAATTG TCATGCCAGT CATGGCATAT TTTTGACCAC	3420
	CTAATTGTGG AATCCAGAAA CTGTGCATTG TATCCATAGC TTGAAGCTTA AATACAACAG	3480
	GACGATCTTT AGGGATTGTT AATGTATTAA CAGTCTCTAT ATGTTTCATCT GGATAAGCAA	3540
20	AGAACCATTT GTATCCTGCA CTTACTGCAT ATACAACCAT TGGATCTTTC TCACTCTTCG	3600
	GTGGTTTTTC GTAATCGTAT AAAGTTTTAA CTGTAGGAAT AGCTAAAGCA GCAACGATTA	3660
25	TGATAGGTAT TACAAACCAT ATTGTTTCAA TGATGGCATT ATGGTGCATC TTACCAGATT	3720
	CGGCATTCTT ATTATACTA TACTTGTAAG TAAAAATGGC GAACATGCCA AGTACAACGA	3780
	AACAAATAAC AAGCATGAAG ACGATTGAAT AAAGAATCAA GAACTTCTGA CTACTTGCTA	3840
30	CTGGCCCTTT TCGGTTGAAA ATTTCTATAT TTGAACAACC ACTAAGTAAA ATTAGTGTGC	3900
	CAATAATAG AAGCAAAGAC TTAAATTTTG ACACTTTTTT GACCTCCTAA TACTACAAAT	3960
	GTAGGGCTTA ACATTAATTT TAAGTTATTA CACAATATTT ACAAGGGCTT ATGGGAAAAA	4020
35	AATTAATAAA ATTGTATCAA AAATGTTGAT AAATCAAGGT GTGACGTGGG TTCACACATT	4080
	TGTTAAAATT ATGTGTACAT TTTGTGACTA ATAGCGTTTT T	4121

(2) INFORMATION FOR SEQ ID NO: 316:

40	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 9310 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
45	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

50	CGAGTGAGTA CAmACATATT TTTATTTGCA AGGGGTAAAT GGCATATAAC TATCTTTTTT	60
	ATGTAAGCTG GTATAAAATT TATACTAATA GGAGGGATAG TATGAATATA GTAGGGCATC	120

55



## EP 0 786 519 A2

	TTGGATTACG	ATTAGTTGAA	AAGTCGGTTA	ATCAAGACAA	TCCTTCAATG	TATCATTTGT	240
	TTTATGGGGA	CGAAgTAGGT	ACAGCCGGAA	CAATTTTAAG	CTTTTTTGAA	ATTCCCAATG	300
5	CGGGTCATAA	GCAGCCAGGT	ACTGAAACGA	TTTATCGATT	TTCATTATTA	GTACCAAATC	360
	AAGCGGCACT	TCATTATTTT	GAAAAACGTC	TTGAGAATAA	TGGTATTAAG	TCTGAACGTT	420
	TGTACTATCT	TGGACAAGAA	GGTGTGTGCT	TTAAAGATGA	AGACGACTTA	GAAATCATAT	480
10	TGCTTGTTAA	TGATAGTTTT	GAAGTACCAC	ATCAATGGCA	ACATAACGCT	TATAGTGAAA	540
	TACCTCAAGC	ATATCAAATT	TTAGGAATAG	GGCCAGTCGA	ATTAAGAGTT	AGAAATGCAG	600
	CGCGTACGGT	AGAATTTTIG	GAAAATGTCT	TAGGTTATCG	CAAAAGAGAT	AATAAATCAT	660
15	TCGATGTGCT	GACATTAGCA	CCACAAGGTT	TATATTCGGA	TTTTGTAGTT	ATTGAGCAAC	720
	AGGGACAACG	TGAAAGACCT	GGACGAGGTT	ATATCCATCA	TATTGCAGTT	AATACACCAC	780
	AAATGAGTGA	CTTAGATGCA	ATTTACAAGA	AATTACAACA	ACAACCACAA	AGTAATTCAG	840
20	GTATAATTGA	TCGCTATTTC	TTTAAATCAT	TATACTATCG	CCATAATTCA	ATTATGTATG	900
	AATTTGCGAC	TGAAGCGCCT	GGATTACTA	TTGATACACC	TGTTGAACAA	TTAGGAAGTC	960
25	AATTGAACTT	GCCTGACTTT	TTAGAAGCAG	AACGTGAACA	AATTGAAAGT	AAGTTACACG	1020
	AAATATAAAG	GAGAATGTTT	AATGGCCAAA	TTAGAAATGA	ATAAAAATAC	GCCTCTTGAG	1080
	TTTGGTTTGT	ATTCCTTAGG	TGATCATTTA	TTGAATCCAT	TGAAAGGTGA	AAAAGTTAGT	1140
30	TATGAGCAAC	GTATTAATGA	AATTATTGAA	GCAAGTAAAT	TAGCAGATGA	AGCAGGTATT	1200
	GATGTTTTTG	CAGTTGGTGA	AAGTCATCAG	GAGCATTTTA	CAACACAGGC	ACATACGGTT	1260
	GTGTTAGGTG	CAATTGCCCA	AGCGACAAAG	CATATTAAAG	TTCAAGTTC	TTCAACGATT	1320
35	ATTAGTGCAC	AGATCCTGTA	AGAGTATTTG	AAGACTTCGC	GACATTAGAT	TIGATTTCTC	1380
	ATGGTAGAGC	CGAAATTGTA	GCTGGCAGAG	CATCAAGAAC	AGGTATTTTT	GACTTGTTTG	1440
	GCTATGATTT	AAAAGACTAT	GATGAATTGT	TTGAAGAAAA	ATTAGGTTTA	CTTTTAGAGT	1500
40	TAAATAAAAC	TGAGCGTATT	ACTTGGTCTG	GAAAATATCG	TCCAGAACTT	AGAAATATGA	1560
	AAATATTCCC	AAGACCAATC	GATAATATAT	TGCCAATATG	GCGTGCTGTT	GGTGGTCCAC	1620
45	CTGCAAGTGC	TATTAAAGCG	GGAAAACAAG	GTGTGCCAAT	GATGATTACA	ACCCTTGGTG	1680
	GCCCAGCAAT	GAACTTTAAA	GGTTCTATAG	ATGCTTATCG	TCAAGCGGCA	ACTGAAGCAG	1740
	GTTTCGATgc	TTCGCCTAAG	TCTTTACCAG	TAAGTACAGC	GAGTCTGTTT	TATACAGCTG	1800
50	AAACAACTCA	GGATGCTATG	AGAGAATTTT	ATCCACATTT	GAATACAGGG	ATGTCATTTA	1860
	TTCGTGGTGT	TGGTTATCCG	AAACAGCAAT	TTGCTAATTC	GTCAGATTAT	CGAGAAGCGC	1920

55

## EP 0 786 519 A2

	GTCATCAACG TTTTATGGCA CAGCTTGATT TTGGCGGTGT GCCATTGAA AATGTTATGA	2040
	AGAATATTGA GTTAATTGGC AACGACATTA TACCGGCGAT TAAAAAGCAT TTATCAAAAT	2100
5	AGGAGGGGCG TCATCATGAA TATTGTATTA TTGTCAGGTT CCACAGTAGG TTCTAAAACG	2160
	AGAATTGCTA TGGATGATTT AAAAAATGAA CTAGAAGTCA TCAATGAGGG ACATCAAATA	2220
	GAGTTGATGG ATTTACGAGA ACTTGAATTA GAATTTAGCG TTGGAAAGAA TTATCTAGAT	2280
10	ACTACAGGAG ATGTATATAA ATTAACGACG TCGTTAATGC AGGCTGATGT GATTTTTATT	2340
	GGTTTTCCAA TTTTCAAGC TTCCATCCCT GGTGCTTTGA AAAATGTGTT TGATCTACTT	2400
	CCAGTCAATG CGTTTCGTGA CAAGGTAATA GGACTTGTAG CGACAGCAGG TTCTAGTAAA	2460
15	CATTATTTAA TTCCTGAAAT GCATTTAAAA CCAATATTGA GTTACATGAA AGCACATACG	2520
	ATGCAAACGT ATGTATTTAT TGAAGAGAAA GATTTTTCAA ATCAACAAAT TGTCAATGAT	2580
20	GATGTTGTAT TTCGGTTAAA AGCGTTGGCA CAATCCACAA TGCGAACTGC CAAAGTACAA	2640
	CAACAAGTGT TTGAAGAAGA AAACAACCAA TACGACTTTT AAAGTATAAA AATAAGACGC	2700
	TCGGCACACT AAATTTGTAA GTGTTTGAGC GTCTTTTCAT ATTAECTATA TAGCCAATGA	2760
25	ACGACGATAA AGGCAAGTGA TGACAAGCAT ATTGAGGTAA TAATGATTGT CATAAGCGGT	2820
	TTAAGTGCGC GATTTTTAAG ATCTTTAAAT GCAACATTTA ACCCTAAAGC AACCATGGCC	2880
	ATTAATAAGC AAATTGTTGA TACAGTATTT AAAATATTTA GCAATGCTGA CGGAATAGTT	2940
30	ACATATGTAT TCACTAAGGC CATAATGACA AATCCAATTA AAAAGTATGG AATGCTTATT	3000
	CGACCCTTGC TAGATGATTC TGATGAACGG AAACGCATAA TTAAAATAAG TACGATGGTT	3060
	AATGGAATCA GTAAGAATAC TCTACCAAGT TTACCAAGAA GTGCAATTTT AAGTGCATCA	3120
35	CTACCACCAA AGCCACCAGC TAAGACAACG TGTGCAATTT CATGAAGACT AACACCAGAC	3180
	CAAGCGCCAT AAACATTTGT CGTCATTGAA AAGATAGCGT AGATAGCTGT ATATATAAGT	3240
	GAAAATATCG TACCAATCAA TGCGATGATA CCGATACTAA TAGCTGTATC CTTTTACGT	3300
40	GATTTGAATA TTGGAGCGAC TGCGGCAATA GCAGCAGCAC CACAAACGCC TGTGCCGACA	3360
	CCTAGTAATA ATGCGATGTT TTTGTCACCA TGCAACAGTT TGTGACAAA GAGCATCATT	3420
45	ACAATACTGA AAATAACGAC ACCTACATCG ATGGCTAATA GTTTACTACC TTGACCGATA	3480
	ATATCGAATA TATTGAGTTT AAGTCCATAT AGGATGATTG CAAATCTTAA TAAATATTTA	3540
	GATGAAAACG TAATACCTGA GCTATATTGT TCAGGATATC CTCTAAAGTG ACGATATAGA	3600
50	ATAGCGATTA ATATCGCGAT AGTTAATGCG CCAACCTTAT CTAGGATTGG CAATTTAGCT	3660
	GCTAAAAAGC TAAATAATGC GACTATAAAT GTTAGTGATA GCCCAATCAT AAAATGCTTA	3720

55

	ATTTTAAAT ATAAATTGG AATGAATAAT AAAGTAGTGA TTAAATTAAG TTGTGTGATA	3840
	GGAAACTTGG ACATCAATCA AAGTAATAGG CACTACAACG CTTATTGGCG GGGCCCCAAC	3900
5	AAAGAAGCTG ACGAAAAGTC agCTTGcAAT AATGTGCAAG TTGGGGATGG GCCCCAACAT	3960
	AGAGAAATTG GGTCCGTAAT TTCTACAGAC AATGCAAGTT GGCGGGGCC CAACATAGAG	4020
	AATTTGAAA AGAAATTCTA CAAGCAATGC AAGTTGGGA AGGACAACAA ATTTAAGATA	4080
10	CAATGCGTAA CATTAAATATG TTATTATAAT GATAATTTAC AGAATTATAT GAAAAATGAA	4140
	TGAGGATGTG ATGGTATGTT TGAATGAAA GTGAATGAAC AAATAACATT AAAAATTTTA	4200
	GAAGCTCATG ACACAGAAGC GCTTTTCAAT TTAGTCAATC GTTCAAGAAA TTCACCTAGG	4260
15	GAATGGTTAC CTTGGGTAGA TGCAACTGAG CAACCATCAG ATACGCGTGC ATTTATTAAA	4320
	AGAGGACTTT TGCAATTTGC TGATGGTAAT GGATTTCAAGT GTGGCATTG GTATGAAGGA	4380
20	ACGCTAGTTG GTGTCATCGG TTTACATGAA ATTAATCACA TGCACAGAAA AACTTCATTA	4440
	GGGTACTATT TAGATAAAGA ATTTGAGGGT CATGGGATTA TGACACAAGC AGTTGAGGCA	4500
	TTGATAAAGT ATGTTTTCGA AGAGCTTGAC TTAAACCGAA TTGAGATTAG TGCCGCAGTT	4560
25	AATAATGAAA AAAGCCGGGC TATTCCTGAA AGGCTGGGAT TTAGTAGAGA AGGTATGTTA	4620
	CGTGACAATG AATTACTAAA TGGTATTTAT TCATCGAGTT ACATCTATAG TTTATTAAAA	4680
	TCAGAAATCG ACCAAAAATG ACAAATTAGA CTTACAAAAG AGTGATGACA TTTAAATGG	4740
30	CAGCGCTCTT TTATTTAATT TTTGAAAATA AAAGGTTGTT GACAGTATTA TTTATAACA	4800
	ATATAATGAT TTTGATAATT ATTATCAACT AGATGATGTT TATGGGAGGA TGCTTTAAAA	4860
	CAGCCGTTTT AAGTGTAATG TATTATTTTA GCGTGTAGG AATGCGAAAA TAATATTTAT	4920
35	AAGAACACAT CTATGGGGAT AATAGAATTT CTATAATGAG GTGTCAAAT GAAAAAGTTA	4980
	ACAACGCTAT TATTAGCATC AACGTTATTA ATTGCTGCAT GTGGGAACGA CGATAGTAAG	5040
40	AAGGATGATT CAAAGACATC GAAAAAAGAT GATGGTGTTA AAGCAGAATT AAAACAAGCA	5100
	ACAAAAGCAT ATGATAAATA TACTGATGAA CAGTTAAATG AATTTTTTAAA AGGTACAGAA	5160
	AAATTTGTTA AAGCGATTGA AAATAATGAT ATGGCCCAAG CAAAAGCGTT ATATCCAAAA	5220
45	GTTCGTATGT ATTATGAACG CTCTGAACCA GTTGCAGAAG CATTTGGAGA TTTAGATCCT	5280
	AAAATTGATG CACGTCTTGC AGATATGAAA GAAGAGAAAA AGGAAAAAGA ATGGTCAGGA	5340
	TATCATAAGA TTGAAAAAGC ATTATACGAA GATAAGAAAA TTGATGATGT GACTAAAAAA	5400
50	GATGCACAAC AATTATTGAA AGATGCAAAA GAATTGCATG CCAAAGCTGA TACATTAGAT	5460
	ATCACACCAA AATTAATGTT ACAAGGTTCT GTTGACCTAT TAAATGAAGT TGCAACTTCT	5520

55

## EP 0 786 519 A2

	GTTGAAGGCG CACAAAAAAT TTATGACTTA TTTAAACCTA TTTTAGAGAA AAAAGATAAA	5640
	AAATTAAGTG ATGATATCCA AATGAAC TTC GATAAAGTGA ATCAATTATT GGATAAATAT	5700
5	AAAGATAACA ACGGCGGTTA TGAGTCATTT GAAAAAGTAT CGAAGAAAGA CCGTAAAGCA	5760
	TTTGCGGATG CTGTTAATGC ATTAGGAGAG CCACTAAGTA AAATGGCTGT GATTACTGAA	5820
	TGACAAATTA TGAACAAGTT AACGATAGTA CGCAATTTTC AAGACGTACA TTTTGTAAAA	5880
10	TGTTAGGTAT TGGCGGTGCC GGTGTTGCAA TTGGCGCAAG TGGTGTGGT AGCATGTGGT	5940
	CTTTCAAATC AATGTTCAAT ACACCAGAAG ATCCGGAAAA AGATGCGTAT GAATTTTATG	6000
	GTAAAGTGCA ACCAGGCATT ACCACACCCA CGCAAAAAAC ATGCAATTTT GTTGC GTTAG	6060
15	ATTTGAAGTC AAAAGATAGA GATGCAATTA AGGCAATGTT TAAAAAGTGG ACGGTTATGG	6120
	CTGATCGTAT GATGGATGGT GATACAGTTG GCAAGCCGAG TAACAATCCT TTAATGCCAC	6180
	CAGTAGATAC CGGTGAATCG ATAGGATTAG GTGCAAGCAA GTTAACGATT ACCTTTGGGA	6240
20	TTAGTAAGTC TTTGATGAAG AAAATTGGGT TATCTAGTAA AATCCCGAT GCCTTTAAAG	6300
	ATTTACCGCA TTTTCCGAAT GATCAGTTAA TAGACGATTA CAGCGATGGT GATATTATGA	6360
25	TTCAAGCATG CTCAAATGAT TCGCAAGTAT CCTTTCATGC GGTTCATAAT TTAGTTCGTC	6420
	CATTTGAGTA TATTGTTAAG GTACGTTGGG CGCAATCTGG TTTTATCTCT GCTAAAGGTA	6480
	AGGAAACACC TAGAAATTTA ATGGCATTTA AAGATGGAAC AATTAATCCT AGGAAGAATA	6540
30	ATCAACTTAA AGATTATGTG TTTATTGATG ACGGATGGGC GAAACATGGA ACTTATTGCG	6600
	TTGTCAGACG TATTCAAATA CACATTGAAA CGTGGGATCG TACTGCGCTG GAAGAACAAG	6660
	AGGCTACATT TGGTCGGAAA CGACATAGTG GTGCACCGTT AACAGGTGGG AAAGAGTTTG	6720
35	ATGAAATTGA CTTAAAAGCG AAAGATAGTC ATGGCGAGTA TATTATTGAT AAAGATGCCC	6780
	ATACGAGGCT AGCGAAAGAA GCAAATACGT CAATTTTACG TAGAGCCTTT AATTATGTGG	6840
	ATGGTACGGA TGACCGCACA GGTAAC TTCG AACAGGCTT ACTTTTATT GCTTTTCAA	6900
40	AAGCGACAAA ACAATTTATC GATATACAAA ATAATTTAGG TAGTAATGAT AAATTAAATG	6960
	AATATATTAC ACATAGAGGT TCTGCTTCAT TTTTAGTATT ACCAGGTGTT AGTAAGGGAG	7020
45	GATACCTTGG TGAAACATTA TTTGACTAAA TTTGTAGCAA TGCTAATAAC TGCTGCTATG	7080
	GTGTGTAGCT TTGGGTTACT GAAAAGTCAG GCAGCAGAAC AACAAAGTAT TAGTGATGTA	7140
	TATAGTGTGA TAACGGATGC GAAATCTGCA CTTTCTAATA ATTCGATATC GAATGACAA	7200
50	AAGCAGAAAG CAATTGAGCA AGTGGTAAGT GCAGTTAAGA AATTATCGCT TGAAGATAAT	7260
	AGTGAAAGTA ATGCTGTCAA ATCAGATGTG AGAAAGCTTG AAGATGCAAA AGCGAATGAT	7320

55

## EP 0 786 519 A2

	GCTAGTAAAG ATGCGGGTTC TAAAATTAAA CTATTGCAAC AGCAAGTCGA TGCTAAAGAT	7440
	GCTGCGATGA CAAAAGCGAT TAAAGATAAA AATAAAGCGG AATTAGAATC TTTGAACAAT	7500
5	AGTTTGAATC AGATTTGGAC AAGTAATGAA ACAGTGATTG GCAATTATGA CGCAAATCAA	7560
	TATGACAAA TTGAAGTCGC ATTATTACAA CTTAGAATTG CAATTCATAA GTCACCATT	7620
	GATACGGCAA AAGTGTGACA TGCTTGGACA ACTTTTAAAT CAAATATTGA TCATGTCGAT	7680
10	AAAAAAGTA ATACGTCTGC AAATGATCAA TACCATGTAT CACAATTAAA TGATGCGTTA	7740
	GAGAAGGCGA TTAAAGCTAT CGACGACAAT CAATTGTCGG ATGCTGcaTg TGCCTTACA	7800
	CATTTTATAG AAACCTGGCC GTATGTTGAA GGTCAAATTC AAATAAAGA CGGTGCTTTG	7860
15	TATACGAAAA TTGAAGATAA AATACCATAT TATCAAAGTG TATTAGACGA ACATAATAAA	7920
	GCACATGTGA AAGATGGTTT AGTAGATTTA AATAACCAA TTAAAGAGGT TGTGGCCAT	7980
20	AGTTATAGCT TCGTCGATGT GATGATTATC TTTTACGTG AAGGGCTAGA AGTGTGTGA	8040
	ATTGTAATGA CATTGACTAC CATGACGCGT AATGTAAAAG ATAAGAAAGG GACTGCAAGT	8100
	GTGATTGGTG GTGCAATTGC CGGACTTGTA CTGAGTATTA TCTTAGCAAT TACGTTTGTA	8160
25	GAACTTTAG GGAATAGTGG CATTCTTCGT GAAAGTATGG AAGCGGGATT AGGTATCGTT	8220
	GCGGTCATAT TAATGTTTAT CGTTGGTGTT TGGATGCACA AACGTTCAA TGCAAACGT	8280
	TGGAATGACA TGATTAAAA TATGTATGCT AATGCGATTA GTAATGGTAA TTTGGTATTG	8340
30	TTAGCGACGA TTGGTTTAAT ATCTGTGTTG CGTGAAGGTG TCGAGGTTAT CATTTTCTAT	8400
	ATGGGGATGA TAGGTGAGCT AGCGACCAA GATTTTATTA TTGGTATTGC TTTAGCTATC	8460
	GTTATTTTAA TCATCTTGC ATTATTATTT AGATTTATAG TTAAATTAAT ACCTATTTTC	8520
35	TATATATTTA GAGTGTGTC GATCTTTATT TTTATTATGG GATTCAAAAT GCTTGGCGTA	8580
	AGTATTCAAA AGTTACAATT ATTAGGTGCG ATGCCAAGAC ATGTTATTGA AGGATCCCA	8640
	ACGATTAACT GGTGGGCTT TTATCCAAGT TATGAACCAT TGATAGCACA AGGTGCTTAT	8700
40	ATTATGGTAG TTGCTATCTT AATCTTTAAA TTTAAAAAAT AAAAAACAGG CCGAGTGCCT	8760
	GTTTTTTTTG TTGCTATATT GGAAATATTC GGTATTGCAG TATAACGATA ATCACAGCAT	8820
45	TGATTCGTAT AAGGTTAATG TGTGGCGGT TTGCCTCGGC ATGTGAACTT AACGATGAAC	8880
	ATACTGAACT CAAAGAGCAA TATGAGTGGC AATGTGAGTA ATATATTTAA TGTAAATCG	8940
	GGTGGTGCAA TGATACTTGC TAATACAAAG CAAGCGAAAT AAATATATTT ACGrTAATGT	9000
50	TTCAATGATG TGGTATCTAT AAGACCGAAT TTTGCAAGAC CCATAAATAA TATTGGTAAT	9060
	TGAAATAGAA GACCAATGT GAATAACCAA CGTATGAGTT CAATCAAATA TGCTTTAAAG	9120

55

GGAAAGCCAA CATAAAATGC AAAAGCGACG CCAGCACAGA ATAATAACAC GCTGAAAAAA 9240  
 CTATATTTAT AAATAAATTG ACGTTCATTA TTATGTAATC CAGGTGCAAT GAATGCCCCAC 9300  
 5 AATTGATAAA 9310

## (2) INFORMATION FOR SEQ ID NO: 317:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3458 base pairs  
 10 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

ATTAGTATTA CACACTAATG TAAATAGATT GGTGGAGAA GAGATATTTG CTAATAAGTG 60  
 CCTTGCTAAT AATGATGTTT AAATTATGAA TTCAATAAAA AAATTAATTG AAGCTGAGTT 120  
 20 ATTAACAACG ACTAATGATT TTGAAGTTAG TATATATAAA AAGACAAGAC CTGAATTACA 180  
 AAGTATTTTA AAAAGTTTTG GTATAAAAAC AACAGGTAAT AAACCTGACT TAATTAAACG 240  
 TATTGACGAC AATTTTCATA TTATTAATAA CTTAGATTTA CCATATGTAT ATATACCAAT 300  
 25 TAAGAAAGGA GAAGAAATTT TAAAGAAAAC AGAGTACTTA ACCTCTTTTA TACAAAGTTA 360  
 TGGTGAAATT TCTCTTGAGC GTGCTTATTA TTTGGTTGAA AACTATATAG ATGAAAATTG 420  
 CGATGATAAA GTTGACAGAA TATACAAGTT TGAATTTCAA AGAAAATATG ACAATGGCGA 480  
 GTTGATTTTT AATCATGGAT ATAATTTCTG ATTGAATATG TTGATAGATC ACTATAAAAG 540  
 AGATGTAAAA GACTACGATA ATGCCAGAAA GTATTCAAAT ATTTATCTTT ACTTTGGTTT 600  
 35 GAGAGATTTT TTAATAAAAT TAATGAGTAA TTATTCATAT TACGATAgTA AAGGGAATAT 660  
 AGATTTGAAC GAAATACAAA ACGATCTGAA TAGATTTATC AACTCTAGCG CTTCTGGTAT 720  
 GTACGAGCGA TTAATATATA ATGAAAATTT ATCCAATAAT ATTATGTTTG AATTATTTAA 780  
 40 AAAGGACACA CAAGATTATA GTGATTTGGA AGAACAAATG ATTGAAAAGT TCATAAACTA 840  
 TGTAGTGTCT AATGTAAAAA AAGAAAGTAG GAGTAATACT CTTATAGAGT TATCAAAAAT 900  
 TTTAGAGAAC GGATATACAA TTGATAAAGA AGAATTTAAA AAAGAAGATG ATTATCTTTC 960  
 45 TAAGTACATA TTTACTGACA TAGATTATTT GAAAAAGTTA GAATCAAAAA TAAACGTTGC 1020  
 TATTGATATT CGAAGTGGAG AAATTCATTT GGTATTAGAT GATGATAGCC TTGATATATT 1080  
 AATACAAAAT CAAAAATACG GCAATGAGTT TTGAGTCATA ACTAAATATA ATATGTTGAA 1140  
 50 GAGAGGTTTT ATTTAAATGG CTAAAATTGG TTATGATCGT GTATCAACGA AAGATTAATA 1200

55

	AACGTACAGA GCTTGTTAAG TGTTTAGATT ATTTACGAGA GAGCGACACA TTAGTTGTCT	1320
	ATCAACTTGA TCGGTTAGGT AGAACGACAA AACTATTaAT TGAATTaTCA CAATGATTCTG	1380
5	ATGATAACGG AATTGACTTA CAAATTAGTA ACATGAACAT TTCAACGAAA GACACAATGG	1440
	GCAAAATGTT TTTTACGATG ATGAGTGCAT TTTTCaGGAT TAGAAGTTAA TTTACTATGT	1500
	GAGTGTATAA AATAGACTTA GCAGCAACAA GAGCGAGAGG CCGAAAAAGC GGGCGCCCCT	1560
10	CTTTACCAGA GAATAAAAAA TGAGAAATTA AATTTTATA TGATGAACAA ACGATAACAG	1620
	GGGAAGAAAT AGCTAGTTAG ACAAGGGTAT GTCACTCAAC TGTTTATCGA GTTGATTAAG	1680
	AAATGAAAAA ACTTATACTA TGAATTACTG TTTAAAAGTG TGCATGTTAT AATATTTATT	1740
15	GAGCAAGTTG GATAGATGGT GGCTAATCTC TTAATAAAGG GGTGATGCCT ATGTTATAG	1800
	TTGTTACTCC TAGnAAAGGA CTAGCATGTC TGATTTTGAA ATGCTTATGG TTGTATTAAC	1860
20	AATCATTGGT TTAGTATTGA TTAGTACTCA AGACCATAAA AAATAaCCTT CTATTCGCTT	1920
	TGACCGGCAT TTTTGAAGGC TATTTTTAAA TAAAATATAA GGTCAACGTC TTTTAAACGG	1980
	GCTCATTAGG GTAACATGTT TCCGAGTGTT GCCCTTTTTG TGTTCaAGA GTTAATGATT	2040
25	TTCATCTTTT GCTTCTACTT GCTACAAATA TATTTTAAAC CATTCTTCTT ATGAATTGTA	2100
	GTTCTGAACA TAATCAGAAT TAATAAAACC AACTTTCCAT ACAGCAGAAA ATACAATTAA	2160
	AAGTATAGAA TGTAATCAGG AATTATATAA AAAGTATTGG AGACCTCTTC ATATATAGAT	2220
30	AATTCACTTA GTTATTTTAG AAAGAAGCCC CTaCAACTA AAGTTGAAA ATAGAGGAAC	2280
	ACAGTTgGAT TACGCATCAA CTGCATaAg CCCCTAAmAA CTAAAGTTGT AAGGGGCyCT	2340
	AAAATTTATT TTGGTTGATT GTCTTCTGGT TTATCTGAAG TCATTGTTTT TGTGTATTA	2400
35	TCATTTAATG AATCTTGTGC TTTTCTGTT TTTGTTTCAA CAGATGTTGT CACTTTATCA	2460
	TTTTCTACTT TTGTATTTTT TGGTTCAGTG ACAACTTTTT TATTCTTGTC AGCTGTTTTT	2520
40	ACTTTATCTT TAGTAATTTT TTGACCACGT TTTAAGTAAT ATTGAACGAT ACCCATTA	2580
	ATGATTGCAT GAATAACAAC GAATAAAATA ATTGTCATTA CTGTATAAAC TCTTATGATA	2640
	TTTTCTGCAA CACTTTGAGA GAGTGATTGC GTATTAAATG AAATCAAGTA ACCAACAGGT	2700
45	GTTTTAAATA AAATAATAAT TAAGTTAAT AATAGTATTC CAATGAAAAA TTTGAAAAAT	2760
	GTTTTTTGAC CATTTTTCAT TGCTTTAAAT CCATTAGCTA AGTGTGTTTT TACTTTATCG	2820
	TTAGTTGATT CaACGAAACT AGTCaTAAAG TTGAAAATAG GTATTAAcAA TAACCAAGTA	2880
50	ATAATTGCTT TAATCAATAA AACAACAATG ATTATGATGC TTTGAGTTGT AATAGCAATA	2940
	CCAATCAGGT GTGTACTATC TGCAATGAT GATTGTACTG AATTCATAAT CATTTCTGAT	3000

55

ATAGATACAA GACCGATCAG AACGCTCTTT AAATAATTAC CTTTTTTAAA TGCAATAAAT 3120  
 AAATCGGTAA ATTTTACTTT TTCATGACTC ATTGCTCGTT TCATAACATT TGTAATTCCG 3180  
 5 ATAAAAATTT GAACTAAAAC AAACAGTGAA ACTACAGCTG CTATAAGTAA TACAGCGATA 3240  
 ACTTTTAAAT ATGCATCTAC TGGTGGTTTT TGCCCAAATT GTGAATAAAT TGCAATAGTT 3300  
 TGTGCATACT TAGCAAGCTG GAAATTAGCA AGTAAATATA CAACTGCAAT TACAGCAAAT 3360  
 10 GCTATAAGTG CATATATCAA TGTATTAAAT AATTGTGGTT TAGCATTTTT AAATGCTGAT 3420  
 TTAAAGTAAG TAAACAAAGT GGTGCCTCCT TTTCTTCA 3458

15 (2) INFORMATION FOR SEQ ID NO: 318:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 695 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

25 GCTTGCAAGA TGCTTTCATT AAAGACAAAG TCATCGATAT TATGAATATG TTCCAAAATC 60  
 AACTGATAT CACTTATACG TTGAATAAGT CTCATGCACA TATTATATGT ACGCCAGAAa 120  
 TATTTGCTAA ATTGTTACAT ACGATTGCAA CTAGAAATAT CGACATTCTA TCTGCCAATT 180  
 30 ATAGATCGTC ATCTATGAGC AAAGCTCGTA TATCATAATA AAAGACACAT CTATACGATG 240  
 ATCATTTACA ATGATTGCGT ATAGTGATGT GTCTTTTTTT GTATTATTTA TCTTCGGACG 300  
 ATTGATCCTC AACCATTGTG TCAAATGCTT CACGAACTTG TGGTACTGTC ATACCTACTA 360  
 35 CAACTTGGAT ACTCTTACCA CTCTTTACTA ATCCGTGTGC CATTTGTTGA TCGTGAAAT 420  
 ATTCCGTATC TGCAACTTTA CTTTCATCAT AAACAGTTAA GCGTAATCTT GTTGACAGT 480  
 TAGTAACATC TTTGATATTT TCTTTGCCGC CTAAACCGTC GAGGTAATAT GCAGCTTTGT 540  
 40 CTTCGTATTC ATTACCAGAT GAAGCACGTT TAGAATCTAC GCTATCACCT TTTTATTTT 600  
 TGTAATCTTG CTTAGAAAAT AATTAACTT CCTCTTCTGT TTyCTTACGT CCAGGTAATG 660  
 45 GAATATCAAA TTyCAAAATT AAGAATCTGA ATAAG 695

(2) INFORMATION FOR SEQ ID NO: 319:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 875 base pairs
  - 50 (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

55



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

5 AATATTGaTa TTATATATAG TCGAATCaAT GATGACCCaC ATCGACCTaA TGCAAATAAT 60  
 AAGACAATCA AGCAGTTAAA TGACTTGTAT TGCTCGCGGA ATTTAAGATA TAATGAAATC 120  
 GCATTCAAGA AATTGATAG TCATTTGCTA TCAATTCAG AAATATATTA TGAATTGCTA 180  
 AATTTAATAA AAGCGAGTGA TCAGTATTAG AGAGAATAGA GCGTTAAGAC TCTATCGCCG 240  
 10 AAGtGCAAGT AATTATTAC GAAACTCTCA GGcAAAAGGr TAATACTGTA ACGCGTTCCT 300  
 GAATTGGTGA TTTATAACA GGGTAGCGAT TGCTATCCTG TTTTATAAT TTAAAGGGGG 360  
 TATTTCAATG TCAAGTGATT TAAAACAAAC ACCTTTATAT CAAAATTATG TTGATAGAGG 420  
 15 TGCAAAAATT GTGGAATTCG GAGGATGGGC GATGCCTGTT CAATTTTCAA GTATTAAAGA 480  
 GGAGCATAAT GCTGTTTCGAT ACGAAATTGG CCTGTTTGAT GTTAGTCATA TGGGTGAAAT 540  
 20 TGAAGTAACA GGTAAAGATG CTAGTCAGTT TGTGCAATAT TTATTATCAA ATGATACTGA 600  
 TAATTTAACT ACTTCAAAG CATTATATAC TGCTTTATGT AATGAAGAAG GCGGTATTAT 660  
 TGATGATTTA GTAATATATA AATTAGCTGA CGACAATTAT TTATTAGTTG TTAATGCTGC 720  
 25 TAATACTGAA AAAGATTTTA ATTGGATTTT AAAACACAAA GAGAAATTG ATGTTGAAGT 780  
 ACAAATGTA TCAACCAAT ATGGTCAATT AGCAATACAA GGACCAAAAG CtAGAGATTT 840  
 AATTAATCAA TTAGTTGATG AAGAkGTAAC TGAAA 875

## (2) INFORMATION FOR SEQ ID NO: 320:

## (i) SEQUENCE CHARACTERISTICS:

- 35 (A) LENGTH: 5897 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

40 TCTTTTTACG AAGGCTACCG TTTAGGATCT GATAACTTAA CTGTAGGAGA AATTGTATTT 60  
 AATACAGCGA TGACAGGTTA TCAAGAACT ATTTAGATC CATCATATAC AGGTCAGATC 120  
 45 ATTACTTTTA CGTATCCATT AATCGGTAAT TATGGTATCA ATAGAGACGA TTTTGAATCA 180  
 TTAGTACCTA CATTAAACGG TATTGTAGTG AAAGAAGCGA GTGCGCATCC AAGTAATTTT 240  
 AGACAGCAAA AGACACTTCA TGACGTTTTA GAATTGCATC AAATTCCAGG GATTGCAGGT 300  
 50 GTTGATACAA GAAGTATTAC GCGTAAAATT CGACAACACG GTGTGTTAAA AGCTGGTTTT 360  
 ACTGATCGAA AAGAAGATAT TGATCAACTT GTCAAACATT TACAACAAGT AGAATTACCT 420

55

	AGTGTCGTAC TTGTAGACTT TGGTAAGAAG CAAAATATTG TTCGAGAATT AAACGTCAGA	540
	GGTTGTAACG TCACAGTTGT ACCATATACA ACTACTGCCG AAGAAATTTT AGCAATGGCT	600
5	CCAGATGGCG TTATGCTATC AAACGGACCA GGTAATCCTG AAGTTGTAGA ATGTGCGATT	660
	CCAATGATT C AAGGAATTTT AGGGAATAAT CCGTTCCTTG GTATCTGTCT AGGACATCAA	720
	CTTTTGGCAT TATCTCAAGG AGCAAGCTCA TTAAATGA AGTTTGGTCA TCGTGGTGCG	780
10	AACCATCCAG TTAATAATTT AGAGACTGGA AAAGTTGATA TTACGAGTCA AAACCATGGA	840
	TATGCAATAG ATATAGATT C GTTAAAAAGT ACTGATTTAG AAGTTACTCA TCTTGCAATTA	900
15	AATGATGGTA CTGTAGAAGG TTTAAACAT AAAACATTAC CAGCATTTTC TGTTCATAC	960
	CATCCTGAAG CAAATCCAGG ACCGTCAGAT TCAAACATC TATTTGATGA TTTTGTAGCA	1020
	ATGATGACTA ATTTTAAGGA AAAGGAGCGT CATATCAATG CCTAAACGTA ATGATATCAA	1080
20	AACAATTTTA GTAATAGGGT CTGGGCAAT TATCATAGGT CAAGCAGCTG AATTGATTA	1140
	TGCTGGAACA CAAGCATGTC TAGCTTTAAA AGAAGAGGGA TATCGAGTTA TTCTGTAAA	1200
	TTCAAATCCA GCGACAATCA TGA CTGATAA GGAAATTGCG GATAAAGTAT ATATCGAACC	1260
25	GTAACTCAT GATTTTATAG CGCGAATTAT ACGTAAAGAG CAACCTGACG CTTTACTTCC	1320
	AACTTTAGGT GGTCAAACAG GTTTAAACAT GCGGATTCAA CTACACGAAA GTGGTGTGCT	1380
	TCAAGATAAT AACGTCCAAT TATTAGGAAC TGAGCTAACA TCAATTCAAC AAGCAGAAGA	1440
30	CCGTGAAATG TTTAGAACAT TAATGAATGA TTTAAACGTT CCTGTACCAG AGAGTGACAT	1500
	TGTAAATACA GTAGAGCAAG CCTTTAAATT CAAAGAGCAA GTGGGATACC CGCTAATTGT	1560
	TAGACCGGCA TTTACGATGG GTGGTACCGG AGCGGTATT TGTATAATG ATGAAGAATT	1620
35	ACATGAAATC GTCTCAAATG GTCTTCATTA TAGTCCAGCA ACGCAATGTT TATTAGAAAA	1680
	ATCTATCGCA GGTTTTAAAG AAATCGAATA CGAgTAATgC GTGaTAAAA CGATAATGCC	1740
40	ATCGTTGTAT GTAACATGGA AAATATTGAT CCAGTTGGTA TTCATACAGG CGATTCAATT	1800
	GTTGTGGCTC CTAGTCAAAC ATTATCAGAT GTTGAGTATC AAATGTTACG TGATGTTTCA	1860
	TTAAAAGTTA TTCGAGCTTT AGGTATCGAA GGTGGTTGTA ATGTTCAATT AGCATTAGAT	1920
45	CCCCATTCAT TCGATTATTA TATTATAGAA GTAAATCCGC GTGTATCACG TTCATCAGCG	1980
	TTAgCTTCAA AAGCAACAGG ATATCCTATT GCAAATTAG CTGCTAAAAT CGCGGTTGGT	2040
	CTAACATTAG ATGAAATGTT AAATCCAATT ACAGGAACAT CTTATGCAGC GTTTGAACCA	2100
50	ACTTTAGACT ATGTGATTTT AAAAATACCA AGATTTCCTT TTGATAAATT TGAAAAAGGA	2160
	GAACGAGAGC TTGGCACACA AATGAAAGCA ACAGGTGAAG TTATGGCCAT TGGTCGAAC	2220

55

	TTACCAAATG GTGAAAGCTT CGATCTTGAT TATATTAAAG AACGTATTTT ACACCAAGAT	2340
	GATGAACGAT TATTTTTCAT CGGCGAACAA TTAGAAGAGG CACAACATTA GAAGAAATTC	2400
5	ATAATATGAC TCAGATTGAT TACTTCTTCT TACACAAGTT CCAAAACATT ATTGATATTG	2460
	AGCATCAATT AAAAGAGCAT CAAGGTGATT TAGAATATCT TAAATATGCA AAAGATTATG	2520
	GATTTAGTGA TAAAACAATA GCGCATCGCT TTAATATGAC GGAAGAAGAA GTATATCAAT	2580
10	TGCGTATGGA AAATGATATT AAACCTGTTT ACAAGATGGT TGATACTTGC GCAGCTGAAT	2640
	TTGAATCTTC AACACCATAT TATTATGGTA CATACGAAAC TGAAAATGAA TCCATAGTTA	2700
	CTGACAAAGA AAAAATCTTA GTATTAGGCT CTGGACCAAT TCGAATCGGC CAAGGTGTAG	2760
15	AATTGACTA TCGACAGTT CACGCCGTTT GGGCAATTCA AAAAGCAGGG TACGAAGCGA	2820
	TAATTGTGAA TAACAATCCA GAAACAGTTT CAACAGACTT CTCAATTTCT GACAAATTAT	2880
20	ACTTTGAACC TTAACTGAA GAAGATGTGA TGAATATCAT TAATTTAGAA AAACCTAAAG	2940
	GTGTCGTTGT ACAATTTGGA GGACAAACAG CGATTAATTT AGCAGACAAA TTGGCTAAAC	3000
	ATGGTGTTAA AATACTTGGT ACTTCACTAG AAAATCTAAA TCGTGCTGAA GATAGAAAAG	3060
25	AATTTGAAGC ACTATTAAGA AAAATTAACG TGCCACAGCC ACAAGGGAAA ACAGCTACAT	3120
	CACCTGAGGA AGCATTAGCG AATGCTGCAG AAATCGGATA TCCGTTGTA GTAAGACCTT	3180
	CTTATGTATT AGGTGGTCGC GCAATGAAA TTGTAGACAA TGACAAAGAG TTAGAAAAC	3240
30	ATATGACCCA GGCTGTAAAA GCGAGTCCGG AACATCCGGT ACTAGTCGAT AGATATTTAA	3300
	CTGGTAAAGA AATTGAAGTT GATGCGATTT GTGATGGAGA AACGGTCATT ATTCCAGGAA	3360
	TCATGGAACA TATTGAACGT GCTGGTGTGC ATAGTGGTGA CTCAATCGCT GTATATCCAC	3420
35	CACAACTTT GACAGAAGAC GAGTTAGCAA CACTTGAGGA CTATACTATA AAATTAGCTA	3480
	AAGGTTTAAA CATCATTGGC TTAATCAACA TTCAATTCGT TATAGCTCAC GATGGTGTGT	3540
40	ATGTTTTAGA AGTAAATCCA CGTTCTAGTA GAACGGTACC ATTCTTAAGT AAAATTACTG	3600
	ATATTCCAAT GGCACAATTA GCTATGCGAG CAATCATTGG GGAAAACTA ACAGATATGG	3660
	GTTATCAAGA AGGGGTTCAA CCATATGCTG AGGGTGTCTT TGTGAAAGCA CCAGTATTTA	3720
45	GTTTTAATAA ATTGAAAAAT GTTGATATTA CTTTAGGACC TGAAATGAAG TCAACAGGTG	3780
	AAGTGATGGG GAAAGATACT ACATTAGAAA AGGCGTTATT CAAAGGGTTA ACAGGTAGTG	3840
	GCGTTGAAGT TAAAGATCAC GGTACAGTAT TAATGACCGT CAGTGACAAA GATAAAGAGG	3900
50	AAGTTGTTAA ATTGGCACAA CGCTTAAATG AAGTTGGCTA TAAAATTTTA GCAACGTCTG	3960
	GAACAGCTAA TAAATTAGCT GAGTATGACA TACCTGCAGA AGTAGTAGGC AAAATTGGTG	4020

55

	TGACTAAAGG TAAAGAAGTA GAAAGGGATG GCTTCCAAAT TAGACGTACT ACAGTTGAAA	4140
	ATGGTATTCC ATGTTTGACA TCTTTAGATA CAGCTAATGC CTTAACGAAT GTAATTGAAA	4200
5	GTATGACATT TACAATGCGT CAAATGTAAA TCAATCAAAC TGTATCGGTG GGGCTGTAAT	4260
	TAACCATTTA CTAAAGAAG TTTATATTAC AGCCTCATT TTTAATGAA TTTCTTAATA	4320
	TAAAGGGAGA CaTATATGAT GAAAGATTTA CCAATTATTG CATTAGATTT TGAATCAAAA	4380
10	GAAAAAGTAA ATCAATTTTT AGATTTATTT GATGAATCAT TATTCGTAAA AGTAGGTATG	4440
	GAACTTTTTT ATCAAGAAGG TCCTCAATTA ATTAATGAGA TAAAAGAAAG AGGCCATGAT	4500
15	GTATTTTATG ATTTAAACT GCATGATATT CCTAATACAG TTGGTAAGGC GATGGAAGGA	4560
	CTAGCTAAAT TGAATGTTGA TCTGGTAAAT GTTCATGCTG CTGGTGGCGT AAAAATGATG	4620
	TCTGAGGCCA TTAAAGGATT AAGAAAACAT AATCAAGATA CAAAAATTAT TGCAGTAACA	4680
20	CAGCTTACGT CAACAACAGA AGACATGTTA CGACACGAAC AAAATATACA AACATCGATT	4740
	GAAGAGGCCG TTTTAAATTA TGCCAAGTTA GCAAATGCAG CTGGTTTAGA TGGCGTTGTT	4800
	TGTTACCTC TTGAAAGTCG TATGTTGACT GAAAAGTTAG GTACATCATT TTTAAAAGTA	4860
25	ACACCAGGTA TTAGACCTAA AGGTGCATCT CAAAATGACC AACACCGTAT TACGACACCG	4920
	GAAGAAGCAA GACAGCTTGG TTCGACGCAT ATTGTAGTCG GTAGACCGAT TACACAAAGT	4980
	GACAATCCAG TCGAAAGTTA TCATAAAATT AAAGAAAGTT GGTAGTATA ATGGCTAAAG	5040
30	AAATTGCAAA ATCATTATTA GATATTGAAG CTGTAACATT ATCACCAAAT GATTTATATA	5100
	CATGGAGTTC AGGTATTAAA TCACCGATTT ACTGTGATAA CCGTGTACG TTAGGTTATC	5160
	CTTTAGTTCG AGGCGCAATC CGCGATGGTT TAATTAACCTT AATTAAAGAA CACTTTCCTG	5220
35	AAGTAGAAGT TATTTCTGGT ACTGCAACAG CTGnTATTCC ACATGCAGCT TTTATTGCTG	5280
	AAAAATTAAA ATTACCAATG AATTATGTTT GTTCATCAAA TAAGAGTCAT GGTAAGCAAA	5340
40	ATCAAATCGA AGGTGCTAAA AGTGAAGGTA AAAAAGTAGT TGTGATAGAA GATTTAATTT	5400
	CGACAGGGGG ATCTTCAGTC ACAGCAGTTG AAGCCTTAAA ACTAGCAGGT GCAGAAGTAT	5460
	TAGGTGTTGT AGCTATCTTT ACTTACGGTT TGAAAAAAGC AGATGATACA TTTAGCAATA	5520
45	TTCAACTACC TTTTACACT TTAAGTGATT ACAATGAATT AATTGAAGTA GCTGAmAmTG	5580
	AAGGTAAAT TTCTAGTGAA GATATCCAAA CATTAGTTGA ATGGAGAGAC AACTTAGCAT	5640
	AATATAGACA CTAGAAGGAG GAATTCAACA AATGAATGAC AAAACATCTA ATGATTTATA	5700
50	TGGGAAGATA AAACATTGTA ACGAATTTAT CAATCATTCA AATGATTCCA ATCTATCTAG	5760
	TAGTCACGAT GTCGACGAAA GTTCAACGAA GCAAAAACAT ATAAAAATA AAACAACATAT	5820
55		

TTAAAAACAA AAAAGCT

5897

(2) INFORMATION FOR SEQ ID NO: 321:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7965 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

15	TCATTCTAAA TCAACTTATT TTCCATTGCA TAAATTGCTG CTTGTGTACG ATCGCTAACT	60
	TGTAATTTAC TAAATATATG ACTGACATGT GTTTTAATTG TTTTTCAGA TACAAATAAA	120
	GTTTCTGCAA TCTCTTTATT TGTTTTACCT TTAACCATTT CACGTAACAC TTCAATTTCT	180
20	CTCTTTGACA ACTTATTCGT GTAGTGTGGT TTTTGGCTAA CTGTTTCGAA TACATCTTGT	240
	GCCTTAGGAT GTATCATTTT TTCACCGTTC ATAACCTGTC TAATAGTTTC AATTAATTGC	300
	TGAGGCTCAA CGTCTTTCAT TTCATAACCA TCAGCACCTT TATTGATTGC TGAAATTACA	360
25	TGTTTCATCAT CAACATAACT TGTTAATACC AAAACTTTAA TATCCGGATA ATGTGCCTTA	420
	ATATATTCCG TAATTTCAAT ACCATTTCATG CCAGGCATCA CTAAATCTAA TAGCACAATA	480
	TCAGGGTGCT CATGCTCTTT TAAATATTCT AAAAATGTTT CTCCATCTGC AAAGTCTTGT	540
30	AAAACTTCTA TGTTTTCAaT CGTGGaTAAT AAAAATCGCA ATCCTTGtCG CACAATATAA	600
	kGGwCATCTA CTAATATnAC TTTGTTTCATG GGTATCTCC TTAaAtCAAG CTATTTTATA	660
	GGAATTGTGA ATTGTATTTG TGTACCTTTT GTTGGCTGAG AATGAAAGGT CACTTTACCT	720
35	CTTAATAATT TAACTCTTTG TTTTATGTTA TTAATACCGT GTGATGAAGC TATCTGAACA	780
	TTATCGATCT CAAATCCTTG ACCATAATCA ATCACGTCAA TATATAGTAT ATCGTTCATT	840
40	TGTTTTAATG TAAGATCCAT TTTATTCGTA TCAGCATGTT TCTTAACATT ATTAATACAC	900
	TCTTGTAATG CTCTGTATAT GTTTTCTTCG ATTTTCATTAG ATAAATCGAT TAAACCTTCT	960
	ACATTTACAT TTAATTGTAT ATGCATTAAT TTAATATATG CTGTCAAAGC ATGAATTAAA	1020
45	CCTTGCTCAA GTCCAACCTGG CTTAAGTnGC CAAATCAATG CACGCATTTT ATTAACGGCA	1080
	TTTTGACTCG TTeCCTCAAT CGTCTTGAAT GCTTGTTTAG CGATGGATTG GTTTGACATG	1140
	CCATACGCAG CATGTGctGT TAGTTTTACA GAAAATAACA TTTGATTTAC TGAATCATGT	1200
50	AAATCTCTAG CTAAACGATT ACGTTCATTA ATTTTTGCCG CTTCTTTTTT ACGGTCTGTT	1260
	AAATAAATAC GTTTGATGGC TGACCCTAAT TGAAATGCGA CAGACTCTAG CAACTCTAAA	1320

55

TGACCCGATT TAAGTGGTAC CGTTGCATGA TGTGTAATAT TGTCATTTTG GCTAGGAAAT 1440  
 GCTTTAGAGG CTAAGTTAAT ACGAGAACAA TTGACGATAT TCGACGCTTT CATTAGCCTA 1500  
 5 CGTTGATTAA ATGCTTTCAC ACACCAACAA GACCCATCTT TAATATAGTG ACAGTGGTCT 1560  
 GCTGTCAAAG ATTGTGGTAG AGCCACATGT GATACAAGTT CATGTTTCACC TACGCTATTG 1620  
 ATGAAAAATA TCCAGCCTGT CGTGAAATTA CTGCCCTCAA TTAAATATTT TAACGCACCT 1680  
 10 TGGGTCATGC TATACATTTT rGTTTCTTCG TTTAAAAATT CGGCAATCTC TTTTAATAAA 1740  
 GCTAGTCGCG TCCTTTGTTC CATCAAATCG CTCCAATTCA TTTTACGTG TATTAATCTAT 1800  
 TATACATTGA GTTATTATAT TTTTAAATCT TAGACGTAA CATGATAAAA TGGCCTTGAT 1860  
 TACTCAATAG TTATATTTTCG GAGAACTGAT TTGTGATATG ATATTAAAGA CTATAGGAGG 1920  
 ATTTTATGAA ATTTAAAAATA CCAGAAAAC TTAATGACTT AAGTTTACGA GATATTTTCC 1980  
 20 AACAACTTAA GGTACCTAAA AAAGATTTAC ATCATTTAAA TATGTCTAAA GATATTACTA 2040  
 TTAATGATAA ACCTGCGCGA TTAATGGATA AAGTGCATAC TGGCGACGAT GTATTTGTTC 2100  
 CAACCATCGA TGAAAAAAGT AATTATGTTC CAAGTTATCG TTATGCACAA ATTAAATACG 2160  
 25 AAGACGATGA TATGGCAATC GTAATGAAAC CTAAGGTGT TAAGACTCAC CCTAATGATT 2220  
 TAAAAGAAAG CAATACTTTA ATGAATCATG TGATTTACAC TATTGATAGT GACTATGTCG 2280  
 AACCAATTCA TCGACTGGAC CAGGAAACAG TAGGATTATT AATTGTTGCT AAAAATCCTT 2340  
 30 TAATGAAAAA AATTCTTGAT CGCATGTTAG AAGACAATGA TATTACGCGG ATATACAAAG 2400  
 CAAATGTAA GGCACTTTTC CCTTTAAAC CACAAACGAT TGATATGCCA ATTGGTAAAG 2460  
 ATAAATTCCA TTCGAATAAA CGACGTGTGT CTCCTACTGG ACAGCGTGCA ATTACACACA 2520  
 35 TTTTAACTTC AAAAATGATA AAAGAAGCTG TGTGCCmACT TGAAATCAAG TTGGATACTG 2580  
 GACGTACTCA TCAAATWCGT GTGCATTTAG CTGAAATTGG TCACCCTGTT ATTGGTGATC 2640  
 CTTTATATGG TGATTCAACG TTAAGACAAT TAGAACTTGA AAGTTACAAA ATAGAGTTTG 2700  
 40 TGcATCCCTT GACTAAGGAA GTCATTTCCG TTTCTTTGGA TGACTAATTT GATTAGTTTT 2760  
 GCATGATATT tAAACATGCA ATACCGCATT GTAACATAA CAAGTATCAA CtTAAAcGGA 2820  
 TAGATGGAAG ATTATTAATT TTTTCAGATG TTCGGTTTTT TTGTTTTTTA CGATGCTTAG 2880  
 GATTTTATAT TTTGATATTT TAGTAATTAT TCATTTTATA ACATCCTTGG ATAATGACTT 2940  
 GTAGTCTTTT TCAACTGCGT TACGTGTATC TATGGACAAT ACATGACATC ATAAGATTTT 3000  
 50 TATCACAGGT TGTTTGGCCA ATACATGTAC AACAAATTCAT CATATAAAAA ATAGGTTCTA 3060  
 TAATAAAACG GACTCCATGA AAAGTTTTTC CTTTTCATGG CTCTATATCA AATCAGACTT 3120

55

	CCAAACTTTA TTTTATATTA ATATTTAATT AATGAGGATC TACCATATCT TCTGGTTTAA	3240
	TCCATGCTTC AAATTGTTCT TCTGTAACAT ATCCAGTTTG AATTGCAGAT TCTTTTAAAG	3300
5	TTAAACCTTC TTTATGGGCT TTCTTAGCAA TTTGAGCTGC TTTTTCATAA CCAATATGTG	3360
	GATTTAATGC AGTAACTAAC ATTAATGATT GATTTAAATA ATTATCAATA TTCTCTTCGA	3420
10	TTGGTTCAAT GCCCACTGCA CAATTGTTAT TAAATGTTTC CATACCATCA GCTAAAAGAT	3480
	AAATTGATTG TAGTGTATTA TGCATAATAA CTGGTTTATA AACATTCAAT TCAAAGTTAC	3540
	CTTGTGAAC TCGGAACCAA CAACTGTATC ATTACCCATT ACTTGGACTG CnACCATTGT	3600
15	TAACATTTCA CATTGTGTAG GATTAACTTT ACCAGGCATA ATTGATGAAC CTGGTTCATT	3660
	TTCAGGGATA GAAATTTCTG CCAAACCAGC TCGTGGCCCT GAAGCCAACC ATCTCACATC	3720
	ATTAGCAATT TTCATTAAGT CTCCTGCTAA TGCCTTCAAT GTTCCATGCA ATTGAACAAC	3780
20	TTCATCATGC GCTGTAAGTG CGTGAATTT ATTTTCAGAA GATACAAATG GATAACCCGT	3840
	ATTTTCTGAA ATATAATGTG CCACTTTATC ACCAAATTCA GGATGCGCAT TAATACCAGT	3900
	ACCAACAGCC GTACCACCGA TGGCAAGATT TAAAATGTGC TTCTTAGATT CAGATAACAT	3960
25	TGTTTCACAA CGGTCAAGCA TATAACGCCA GCCACTAATC TCTTGTCCCTA GTTGATCGG	4020
	CGTTGCATCT TGTAATGTG TACGACCAAT TTTAATAATT GAATCAAATT TATCTTCTkT	4080
	TTCTTTCAAA GTATTTCTTA AAAGTTTTAA TGCAGGTTCT AATTTTGTTT CAACCTCTTG	4140
30	ATATAATGCA ACGTGCATAG CAGTTGGGaa TGTATCaTTC GaACTTTGAG ATTyATTAC	4200
	ATCATCATTk GGGTGGATAC TTTCATCACT TTGATGATCT TTTAAATACA TATTAGCAAC	4260
35	ATAACTTACT ACTTCGTTCA CATTCAATT ACTTTGTGTA CCGCTTCCTG TTTGCCATAC	4320
	AACTAGTGGG AAGTGTTCAT CTAATTCACC TGATAAAAT TGAATCACATG CGTATACAAT	4380
	GGCATCTTTC TTTGCCTCGC TTAATTTTCC TAAATCAAAA TTAGCTATTG CTGCTGCACG	4440
40	CTTTAGTTGT GCAAAACCAT AACTACTTTC GATTGGCATA CGCTCTTTAC CAACTGGGAA	4500
	ATTACGTTTA CTTCTTTCTG TTTGAGCACC CCAATATTTA TCTGCAGGTA CTTCTATTTT	4560
	TCCAAAAGTA TCATGTTCaA TTCTTACTGA CATTCAATTT CTCCCCTTAT CACTGTTTAT	4620
45	TTAACTGTAG TATATCATTA AATAATTTAA TTGAGCAATT TATGATTAAA ACGTTTTCAT	4680
	AATTTGAAAT AAAAATACAC TAATCGCACG TGTTACCCTT TTATTACAGT GATACGGTCA	4740
	TACGATTAGT GTGTTATCTA TCATTATTTA GTTATTATTG AACTAAGTTT AATTACGATA	4800
50	CTTTGTTTIA GTAGCTTCAA CCGTAGCAAT AGCTGTAAGT ATATATAATA CAGCACTAAC	4860
	AATTGTCGTA TATGGATTIA GAGCAACAAG CGTaCCTAAA ACTCCTGTIA AACTCGCATA	4920
55		

TACAATACCT GATTGATTAC TTTTAATGAA TGTTCGCGCA TTAACATCAT CAATTAATCC 5040  
 TTTTGATAAA TTGAGTTGTA ATTTTATTAC TTTGAAAATA ACAGGTAAAT ATAATGCCCC 5100  
 5 AATTGCCAAT GGAAAAGCTT TAATTGATAT TAAACTTATA ATAAGTGTG CTATCAATAA 5160  
 TTGAATCCAG TATTTTCCTA ACATAAATAT ATAAATCTCC TCTAATTTCA TTCTTCAATA 5220  
 GCATATCATA ATCTTGGCAT ATTAAGAAAC GCGGTTTAAT GATTTTCATTA AAAATATTAC 5280  
 10 TGATAGATGA CTTCTTTCAA TTATGTCTGG AGTAATTAAT TATCAATTCC GTTTAAATGG 5340  
 TGTTTTAATA TTAAAATTG AACTTTTGAT ATATTACTAT GTCTGGTACA CAAATCAATG 5400  
 15 TTTTATGCTT TACAAAGTTA TATTGGCAGT AGTTGACTGC AGTCCACAAC ATAGAGGCTT 5460  
 CGGAATGTCA GCTTCTATTT CATGCAAGTT GGTGGAGCTC CAACATAGTG GAATTGGATT 5520  
 CCCAATTTCT ACAGACATTG CAAATTGGGG AAACGGGCCA CAACTCAGA AACTGGTGGA 5580  
 20 AAGTCAGCTT AAAATAACAT GCAAGTTGGC GAGGCCCCAA AATAGTGAGA TCGGATTTCT 5640  
 AATTTCTACA GACATTGCAA ATTAGGGAAA CGGGCCACAA ACTCAGAAGT TGGTGGAAG 5700  
 TCAGCTTAAA ATAACATGCA AGTTGGCGGG GCCACAACAT AGAAAAATTG GATCCTCAAT 5760  
 25 TTCTACAAAC AATGTAAGTT GGGGAAACAG CCCCACACT GAACTAGCA GAAAGTCAGC 5820  
 TTCTATGAAT ATAATAAAAA AGCTAGGTAA CAAAATGCTA CCTAAGTTCA TATTCAAGAT 5880  
 AATCAATCCT ATTTGATATA TGTTCTATAC TATACATTAT TTACATGATA AATAACTGAA 5940  
 30 TATTACACAA TTATAATACT TTAGTACTG TCTTCTTCAG AATCTTTTC TTGATCATTT 6000  
 TGATCAGAGA TTTGTTCCaT TTCTTTACCT AATTCTTTTA AATCTTCAA ATCCGTTACC 6060  
 ATACTGTTTT CTCTTCATG ATAATTTAAT TTTGGATCTT TGTCTTTAGA CATAATCAAT 6120  
 35 ACCTCACAGA TTTTAAATTA AGCAAAGCGT GACGTAAAGT AAGCTTTAAC ATCTTCAGGT 6180  
 AAACCTGCAG CCGCTTCTTT ATCAAGAATA ACATTTACCA TTCTATGTGC TTTTAAATCG 6240  
 40 GCTGGTTCGA AGCTTGTTTT ACCATTTTCT TGATATAATT TTTCAACTAC ATCTCGTTTA 6300  
 TTAGCACCTG TCACTACTAA GAAAATTTCT CTTGCTTCCA TTAGTCCTTG ACGAATACTA 6360  
 ACATTTAACT TACCTTGCTC ATCGATAGAA ACAACTTGTA ATGTTAATTT CCCTTTATTT 6420  
 45 TCTTTAGTTT TAATCTTATC AGCGATTAAT TCGATTGCAT CTTTTTCATA AGCAATTGGA 6480  
 TAAACTTGAC CTGCTGGTAC ACCTAACGCT TCGAAATATG ATTTTTTATC GTCATAATCT 6540  
 AAAATATTTA TTTGGCTAAA ATCAACAGCA TGTTTTTTCAA CATTTTCTT TAATTCATCT 6600  
 50 AGAACTGGCG CTTGATCTGT ATCTAAATGA AAACCTGCAA TTGTAGTAGG ATTATTGTTA 6660  
 AATTGCTTTC TAATAATATC AGCAGCATAT TCTGCTACAA GTTGACTATT GTCAAAGACT 6720

55



GATATTAATA ATTATTATAC CCTAACTTTC AATATATCAA ACCATTTAAC TTAAACATGC 6840  
 TTATACTCTA AATATAGCAC TTAAGCATCA TTTTATAAT GAAAATGAGT AAATTTTAAT 6900  
 5 TCAATCCCGG TAAATCTTGT TGACGTAACG CTTCATAAAT TAACAACGCA GCAGTATTTG 6960  
 ATAAATTTAA TGAACGAATA TGTTCACTCA TAGGAATTCT TAACGCTGTG TCTTGATATT 7020  
 TCTCTTTCAC CCAGTCTGGT AATCCTGTGC TTTCTTTTCC AAAAATGAAG TAAAAATCTT 7080  
 10 TGTCA TGATT TGA AAAATCA AAATCACTAT AAGTCTTTTT ACCAAATTTT GTTAATAAGT 7140  
 AATACTCGCC ATTTGTGACT TCAAAAATG CTTCAATACT ATCATGATAC GTAATATTCA 7200  
 CAAATCCCA ATAATCTAAA CCGGCTCTTT TTAACATTTT ATCATCAGTT CTAAATCCAA 7260  
 15 GAGGTTTAAT TAAATGTAAA TGTGTGTTTG TACCTGCACA CGTACGgCAA TGTTACCAGT 7320  
 ATTAGCTGGG ATTTCTGGTT GATATAAAAC GATATGATTT GTCATATTAC TATTCTCTCC 7380  
 20 TTGTGTCTAA TCCTTTTATC ATTTCACTCT GAACTTCTGC ATCCTCTTGA TCATAATTAG 7440  
 CATTGATAAA ATCTCTTGCT TCTTCCCCAA GAATTTGACC AATGGCCCAA TAAGCAGTTG 7500  
 CTCGAATCAA CGGTCTTTCA TCTGTTGTTG CAACTTTTTT CAATTCTGGA ATTGCATCCA 7560  
 25 CTTCAATAAA ATGCGCCAAT GCTAAAATAG CATTTCGTTG TATCGGCTTT TTACCACGCC 7620  
 AAGCACCTGC AAGgTGACCA TATGTTTGTT TGAATTCCTT ATTAGACATA CGTAGtAAAG 7680  
 GTACTAATCT TGGCTTTAAA ATTTCTGGTT CAAAATGAT GTCATCTTGT TCGGTATTAA 7740  
 30 TACCTCTATT TTTCGGACAA ACTTGTTGAC ACGTATCGCA ACCATATAAT CTATTCCCAA 7800  
 TTTTATAACG ATATTGGTCA GGCATATAGC CTTTGTGTTG CGTTAAAAAA CTAATGCATT 7860  
 TCTGACTATT TAATTGGCCA TTCCAACTA ATGCACTTGT TGGACAACGA TCAACACAAA 7920  
 35 TTGTACAATG CACCACAGCT ATCTAATAAT GGATCATCAG GTTCC 7965

## (2) INFORMATION FOR SEQ ID NO: 322:

(i) SEQUENCE CHARACTERISTICS:  
 40 (A) LENGTH: 1302 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

GCCCTGTTGG AGAAATCACC TTTATACGAT GGTGAAAAA GCATCATTAG GTACAATCGT 60  
 50 GGCAATTATA GTATTTTGC TGTTTTAAAG AAACATTCGT ACGACGGCAA TTTCTATTAT 120  
 ATCGATTCCG TTATCACTTC TTATGGCGCT TATTGCTCTG AAATTGAGTG ATGTTTCATT 180

55

TGTAGTTGTT GAAAATATTT ATCGACGCTT AACAGATTCA GAAGAACAAC TAAAAGGTGA 300  
 AAATTTAATT ATCAGTGC GA CAACTGAAGT ATTTAAACCA ATAATGTCAT CGACACTAGT 360  
 5 TACTATTATC GTCTTCTTAC CACTTGTGTT TGTATCAGGT TCAGTAGGCG AAATGTTTAG 420  
 ACCTTTTGCA TTGGCTATTG CATTTAGTTT ATTAGCATCG TTATTAGTGT CAATTACACT 480  
 CGTTCCAGCG TTGGCAGCTA CACTATTTAA AAAAGGCGTT AAACGTCGTA ATAAACAACA 540  
 10 TCAAGAAGGA TTAGGTGTTG TTAGTACAAC TTATAAAAAA GTATTACATT GGTCAATAAA 600  
 TCATAAGTGG ATTGTAATTA TATTAAGTAC ATTAATTTTG GTTGCAACTA TTGTATTTGG 660  
 AGGACCGAGA CTAGGCACTA GCTTTATTTT AGCAGGTGAC GATAAATTTT TAGCTATTAC 720  
 15 TTATACACCG AAGCCTGGTG AAACGGAGCA AGCAGTGTG AATCATGCGA AAGATGTTGA 780  
 AAAATATTTA AAACAGAAAA AGCATGTAAA AACAATTCAA TACTCAGTTG GCGGTAGTAG 840  
 TCCAGTAGAT CCAACGGGTA GTACAAATAG TATGGCAATC ATGGTTGAAT ATGATAATGA 900  
 20 CACGCCTAAT TTTGATGTAG AAGCGGATAA GGTTATTAAA CATGCAGATG GCTTTAAACA 960  
 TCCTGGAGAG TGGAAAAATC AAGATTTAGG AACAGGTGCA GGTAATAAAT CTGTAGAGGT 1020  
 25 TACTGTAAAA GGTCCATCAA TGGATGCCAT AAAATCAACT GTAAAAGATA TTGAACAGAA 1080  
 AATGAAACAG GTTAAAGGAC TAGCCAATGT CAAATCTGAT TTATCGCAA CATATGATCA 1140  
 GTATGAAATT AAAGTCGATC AAAATAAAGC GGCAGAAAAT GGTATTTCTG CAAGTCAACT 1200  
 30 TGCAATGCAC TTGAATGAAA ACTTACCAGA AAAACAGTT ACGACTGTTA AAGAAAATGG 1260  
 TAAACTGTT GATGTTAAAG TCAAACAAAA TAAGCAAACA GC 1302

## (2) INFORMATION FOR SEQ ID NO: 323:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1003 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

45 ATATATATTA ATTTAAACGT GTTTCACATG TACCAAGTGT AATGACAGAT AATGCTGCGT 60  
 TTAAACCACC TTCAACAAGG TTTTGTACTG CTTTCATCAGA GAAGAAAGCA ATATGTGGTG 120  
 TTACTAAAAT TCTTTCATGT TCGATTAACT CTAATAATGT TTTATCGTCA ATGTCTTTAT 180  
 50 TAGTCCAGTC ATTTGTGAAG TATGCTGCTT CATTTTCATA AGTATCAATC GCAGCACCTA 240  
 ACAAAGTACC ATCGTTCACT GCAGCGATTA AATCAGGTGT ATTGATGACT GCACCACGTG 300

55

5 GATAGCTTTC TTTGTTGCT GGAACATGTA AAGAAATAAT ATCGGCATCT TTAATAGCTT 420  
 CTTTAACTACT ATCTTTATAA GTTAAAAAGT CTAAATCTTT ATTAGGATAG GCGTCATAAG 480  
 CTGTAATTGT AGCACCAAAT CCTGCATATA TtTTAGCTGT AGCAGCACCG ATACGACCCG 540  
 TACCAATAAT TGCAACAGTC ATATTTTAA CTGGTTTAGA CATGATTCT GCTTGCCAAG 600  
 TAAAATCATG TGCTTGTAACA CGGCGTTCAA TATCTGGGAA GCGACGCACT AATTGTAGGG 660  
 10 CGATAGATAC AGAATACTCT GCAATtGTTT CAGGTGAATA ACTAGGAACG TTAGATATCA 720  
 CAATATTGTG TTTTTTAGCT AAATCTAAAT CATACATATC AAATCCAGCA GTACGTTGTG 780  
 CAATTTGTTT AATACCGTAA GATTCTAATT TAGGATAAAC GTCATTTTCT AACTTACCAA 840  
 ATTGCATTGT AGTTACGCCA TCGTAATCTT TTAATTGATC GACTGTAGCA CTTGATAATA 900  
 GCTCTTTAGA AGTAGTTACT TCGACATTAT TCTtTtTCCC CCAATTTAAT GCCATCTCTT 960  
 20 TCTCATAATC ACGCGTaCCA AAGAACATAA TTCTCGTCAT TAT 1003

## (2) INFORMATION FOR SEQ ID NO: 324:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5030 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:

GGCTTTTATA GTGTTTACGA TTATTATGAT TACTTATATT CACGCTGTTA AAAATTATCC 60  
 35 AAATAACCGT ACTGTTCAAT ATGGTTATAC AGCTGCATTT ATACTTGTTA TATTACAAGT 120  
 TATCACAGGT GCATTATCTA TTATGACAAA TGTTAACCTG ATAATCGCAC TTTCCATGC 180  
 ATTATTTATC ACTTATTTAT TTGGTATGAC AACATACTTT ATCATGCTAA TGTTACGATC 240  
 40 AGTAAGAAGT GACAAGCAAT AACaAAAAAG cmAACCGTAA TTTTAATGGC ACGCCCATTA 300  
 AAATTACGGT nTTTTATATC AATATTTAAA AtTAAACcTA AGCCATGTAA AAACGAGATT 360  
 ACACGTCAAT TGTTGTGTAA TCTCGTTTTA TnTTAATCAT TTTAGTCAGT TGCTTTTTCA 420  
 45 ATTTGATTA ATAAATCGCC TGTCGCTATT GTGTCACCAT TATTTACAGT TACTTGTTTA 480  
 ATCACACCGT CAAATGGTGC TTGAATTGTT GTTTCCATTT TCATAGCTTC AGTAATTAGC 540  
 50 AACGGCTGAT TAGCTTTTAC AGTTTCACCT AACTAACCT TGACTTCAGT TACTGAACCT 600  
 GGCATTTGAG CACCGATATG ACTTGGATTA CTCTTATCTG CTTTGGCTT AACGTTGCA 660  
 TTTGTATGCA CATTTTCATC TTTAATGTAA ATACGTCTCG CTTGACCATT CATCGCATAG 720  
 55

	CGTTTACCTT TATCGATTTC GATTCTACT GTTTCACCAT TACGCATTCC AAAGAAGAAT	840
	GTAGGCGTAT CAAGTAACGA TAAGTTTCCG TATTGATTTC TAGTTTGAAT ATATTGTTCA	900
5	TATACTTTTG GATATAGTAC ATAAC TAATA ATATCTTGCT CCGTAACAGG ACCTTGTTGC	960
	TCTTCTTCAA GCAACTCACG GACTTTTTCA AAATCAACTG GCTCTAGATA TTCACCTGGA	1020
	CGAGCTGTTA GTGCTTCTTG GCCTTTTAAA ATAACCGCTT GTAAATCTTT ATTAAAACCA	1080
10	TTTACAGGTT GTCCTATTTC ACCTTTGAAG AACGACACTA CTGATTCTGG GAAATCTAAT	1140
	TTATAGCCAT CTGTAATCAC GGATTGTTCA TCAAGATCAT TTTGTACCAT ATAAAGTGCC	1200
15	ATATCACCAA CTACTTTAGA CGATGGTGTT ACTTTTACGA TATCACCAA TAAGAAATTC	1260
	ACTCTGCGAT ACATATCTTT GACTTCATCA AATCTTTCGC CTAAACCTAA ACTTTTAGCT	1320
	TGTTGACTTA AATTCGAATA CTGTCCACCA GGCATTTCAT GTTGATAAAT TTCAGTATTC	1380
20	GGTGATTGTA TATCACTTTT AAAGTCTGAA TAATAAGTAC GTACAGTTGA CCAATAATGA	1440
	CTAAGTGA CTATACCTTC AATATCAGTT CTAAGGTGGC GTGGGAAGCC ATTTAATGCA	1500
	TAATATAACG AATTGGCGCT TGGCTGACTT GTTAAACCAC TCATTGAAGC AACAGCAGTA	1560
25	TCAATGATAT CGACACCAGC ATCTATTGCT TGTTTGTATG TTAATAAACC ATTACCACTT	1620
	GTATCATGAG TGTGAAGATG AATTGGTAAA TCTACAGCTG ATTTTAACTC ACCAATCAAT	1680
30	TCGTAAGCGG CTTTAGGTTT TAATAAGCCT GCCATATCTT TAATCGCTAA AATATGGAAA	1740
	CCTTCACGTT CTAACCTTTT AGCTAGTTTG ACATAATACT CTAAAGTATA AATGTTTGAT	1800
	CGCTCAGGAT TTAAAATGTC ACCTGTATAA CAAATAGTAC CTTCTGAGAT TTTGCCCGCT	1860
35	TCTTGTACTG CTTTATTGGC AACTTTCATT TGATCTACCC AGTTTAAATGA ATCGAAAATT	1920
	CTAAAGACAT CTATGCCTGC TTTAGCACTT TCTTGTACGA ATTTATGAAT AACATTATCA	1980
	GGATAGTTTT TATAACCAAC TGCCTTTGAA GCACGTAACA ACATTTGGAA TAATACATTT	2040
40	GGAATAGCTT TACGTAGACG TTCAAGTCGT TCCCATGGGT TTTCCCTCAA GAAATTATAT	2100
	GCCACATCAA ATGTAGCACC GCCCCACATT TCTAGTGAGA AACCATCTTT AAATACGTCC	2160
45	GCTGTTTTGG ATGCGATATT AATCATATCC TTAGTTCTAA CTCGTGTAGC TAATAATGAT	2220
	TGGTGTGCAT CTCTAAAGGT TGTATCTGTT AGTAAGACAT CATCCTGCTT TTTAACCCAT	2280
	TCAGCTACAC CTTTTGGACC TACTTCATCA AGCAATTGTT TCGTACCACT AAATGAAGCG	2340
50	ATTTTACTTG AAGATACAGT TGGAATTGAT GCTAATTCAT AGTCTGGTTT CGGACGTTTC	2400
	TCAACATTTG GGAAACCATT AATTGTTACA TTACCTATAT ATTCTAATGT TTTAGTACCT	2460
55	CTATCTAGAG ACGGCTGAAT GTCGAAAAGT TCTGGTGTTC CTTCAATAAA TTTAGTTGTG	2520

## EP 0 786 519 A2

	TTAACACCAC GAATACGCAT TTCTCGTAAT GAGCGTACCA TTTTCTCTTC TGCTTGTTTA	2640
	AATGATATCG CGTGTGTAGA TAATTTTACG AGTAATGAAT CATAATAAGG TGATATCTCA	2700
5	GCACCTTGGA AACCATCTCC AGCATCAAGA CGTACACCAA AGCCCCGCT TGAACGATAA	2760
	GCAATGATTG TTCCAGTATC CGGCATGAAA TCATTTAACG GATCTTCTGT TGTAATACGA	2820
	CATTGGATGG CATAGCCTAA TGTTGTAATA TCTTTTGTGTT GCGGCATATT AATCTCTTCA	2880
10	CCAAATAAAT CGGCACCTGC TGCAACTAAA ATTTGTGTCT TAACAATATC AATTCCTGTT	2940
	ACCATCTCTG TAATTGTATG CTCTACTTGT ACACGAGGGT TAACTTCTAT AAAGAAGAAT	3000
15	TCGTCACCAG ATACTAGAAA TTCAACAGTA CCTGCATTGA CATATTTAAT ATTTTCCATC	3060
	AATTGAATTG CAGCATCACA AATACGTTGA CGTAATGTTG ATGATAATCC AACTGATGGT	3120
	GCAACTTCTA CAACTTTTGT ATGACGACGT TGTACTGAAC AATCACGTTT AAATAAGTGT	3180
20	ACGATATTTT CATGTTTCGTC ACCTATGACT TGTACTTCAA TATGCTTTGG ATTATCAATG	3240
	TATCTTTCTA TGTAAACTTC ACTATTACCA AATGATTTTT CAGCTTCTGA TTTTGCTCTA	3300
	TGGAAAGCAT CTTCTAATTC ACTTTCTTCA CGAACGATTC TCATACCTTT ACCGCCGCCA	3360
25	CCACTTGTGG CTTTAATCAT TAGCGGGAAA CCAGCTTCTT CTGCAAATTC TTTTGCTAAT	3420
	TCATATGATT TAATTGGACC GTCTGTACCA GGAATAACTG GTAAATCTGC CTTGATAGCC	3480
30	GTTGTACGAG CTTTAACTTT ATCTCCAAAC ATATCTAAAT GTTCTAAATG AGGACCAATA	3540
	AATTTAATTC CTTCTTCTGC ACAACGACGC GCAAATTGTT CATTTTCACT TAAAAATCCA	3600
	TAGCCAGGAT GAATCGCATC CACATTCGCT TGTTTTGCTA CATCAATGAT ACGCTCAATA	3660
35	TTTAAATAAC TTTCAGCAGG ACCTAAATCA CTTCCAATA AATAGGATTC ATCTGCTTTA	3720
	TATCTATGTA ATGAACTTTT GTCTTCATTC GAATAAATTG CAACTGTGCT GATGTCTAAT	3780
	TCTGCCGCCG CTCTGAATAT ACGAATTGCA ATTTCTCCAC GGTTAGCAAC AAGTAACTTT	3840
40	TTTATTTGTT TCAATAGCGA TACACTCCTC AAATATTAG AATTTTCTAA CTAATTAGAT	3900
	AATAAAATTT TATCTTAAAG CGCTCTGTTT TGCTATAGTt mTGTTTCmAA TTTTCAAAaT	3960
45	TTaACATyCT tGAGACAATT AAAaCCyCCG CTTcmGaAAT AATAATTTCA AAAATGACTA	4020
	TGCAACAACA GGTAGTTCCA CGTTTTTGTT GTGAAACATT TTCGATTTCT ACAACTCTAA	4080
	AAAATTAAAA ATAAAAATTGC AAAACATCAA CATTTATTAT CAATAGCGAT AACTTTATCT	4140
50	TATCATCATG ATTCTAATTT CGCCACCACA TTTAGTAATT TTTAGTCATA AAATTTAGTT	4200
	ATAATTATAC GTTGTTTTGT TTATAAAATT TGATAATCak GAGTAATCtC GTAATATCAA	4260
55	AACaAAAAGG AAGTTAAGCG TTGTTTGGTT GcCTAACTTC CGTTATTGAA CTCATCcAGT	4320

TCTCGTACTA AATATTGGCT AGTATTTTTT TAATTAAATT GTCTTCTTAT ATCAACTTTT 4440  
 TGTGTGTGTT TCTTTCGTTG CTGGTCTACT TTGATTGTGTT TACCTACAAT CAGAAGTAAA 4500  
 5 CCCATAGCAA TACTTAAACT AATCATTGAT GATCCACCAA AGCTGATAAA TGGCAATGGC 4560  
 ACACCAGTTA ATGGAATTGT TGCCGAAATA CCGCCAATGT TTACAAACGT TTGACTTCCA 4620  
 AAGTATGTGG CAATCCCAAC ACACACAAGT TTATAAAAAT ATGATGATGT TTTATTTGCA 4680  
 10 AACTGGAAGG CACGATATAC AATAAAGAAC TCTAAAGTAA TAACTAGCAA TCCTCCGATT 4740  
 AAACCTAATT CTTCGCAAAT AATTGCAAAA ATAAATCTG TATGTGGTTC TGGTAAATAG 4800  
 15 CCCAATTTCA TTGCACTATT TCCTAATCCT TTTCCAAATA CGCCACCGTT ACCTATCGCA 4860  
 AGCAATGAAT TGGAAATATG GTATCCAGTT CCTGATTCTGA ATTGGAATGG ATCTGTTAGC 4920  
 GTACTAAATC TGGCAGTTAA ATAACCTGGT AACCAACCAG CCATTAATGC AATGACAAAT 4980  
 20 ACTACTAAGA ATCCTAGCAC TGCTGGTATA CCAAATCTTA GGACTTTGTT 5030

(2) INFORMATION FOR SEQ ID NO: 325:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1389 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

CTTGTTAATC CGAAAATTAT TAGTCAATCA AATGAAACAA TAACAGACTT nGAAGGTTCA 60  
 35 ATTACATTGC CAGATGTTTA CGGCGAAGTG ACAAGAAGTA AAATGATAGT TGTCGAAAGT 120  
 TATGACGTCA ATGGGAACAA AGTTGaACTA ACTGCACATG aAGATGTAGC AAGAATGATT 180  
 TTGCATATTA TAGATCAAAT GAACGGTaTC CCTTTTACAG AACGTGCCGA CCGTATTTTA 240  
 40 ACAGATAAAG AAGTGGAGGC ATATTTTATA AATGACTAAA ATAATATTTA TGGGTACACC 300  
 AGACTTTTCA ACAACTGTTT TAGAAATGCT TATTGCAGAA CATGATGTCA TTGCAGTCGT 360  
 AACGCAACCA GATCGACCTG TTGGACGTAA ACGTGTATG ACACCACCAC CAGTTAAAAA 420  
 45 AGTTGCAATG AAATATGATT TACCTGTATA TCAACCTGAA AAATTAAGTG GATCAGAAGA 480  
 ATTAGAACAA TTGCTTCAAT TAGATGTAGA TTTAATTGTA ACTGCTGCTT TTGGACAATT 540  
 50 ATTACCTGAA TCATTGTTGG CATTACCAAA TCTTGGGGCA ATTAATGTAC ATGCATCATT 600  
 GTTACCGAAG TATAGAGGTG GTGCACCAAT TCATCAGGCA ATTATCGATG GTGAACAAGA 660  
 AACC GG CATA ACAATTATGT ATATGGTTAA AAAATTAGAT GCGGGTAATA TTATTTTCGCA 720

55

ATTAGGGGCA GATTTATTAA AAGAAACTTT ACCATCTATT ATAGAGGGCA CAAATGAAAG 840  
 CGTACCTCAA GATGATACGC AAGCAACATT TGCTTCCAAT ATTCGACGCG AAGATGAGCG 900  
 5 AATTAGCTGG AATAAACCAG GAAGACAAGT GTTTAATCAA ATTCGTGGAT TATCACCATG 960  
 GCCAGTTGCT TATACAAC TAAGATGACAC TAACCTGAAA ATATACGATG CTGAACTCGT 1020  
 TGAGACTAAT AAGATAAACG AGCCTGGAAC CATTATAGAA ACGACTAAAA AAGCCATTAT 1080  
 10 TGTGCTACA AATGATAATG AAGCTGTTGC AATTAAAGAT ATGCAATTAG CTGGGAAAAA 1140  
 GAGAATGTTA GCTGCCAATT ATTTAAGTGG TGCGCAAAAC AACTAGTAG GGAAGAACT 1200  
 TATATGATAG AAAACGTGAG AAGTCTTGCT TTTGACACGA TTCAAGATAT ATTAAATGAA 1260  
 15 GGTGCGTATA GTAACCTGCG TATCAATGAA GTGTTGTCAG AAAATGAATT AAATGCAATG 1320  
 GATAAGGCTT TATTTACAGA AATTGTCTAC GGAACCGTTA AAAGAAAATA TACGTTAGAT 1380  
 20 TTTTATTTA 1389

(2) INFORMATION FOR SEQ ID NO: 326:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2746 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

TTTGCTAATA ACAATAAAGC CAAAGCCGAT TCACACTCTA AACAGCTAGA AATTAATGTT 60  
 35 AAGAGTGACA AAGTACCTCA AAAAGTAAAA GATCTAGCAC AACACAATT TGCTGGTTAT 120  
 GCTAAAGCAT TAGATAAACA AAGTAATGCA AAAACTGGTA AATATGAACT TGGCGAaCTT 180  
 TTaAAATTTA TAAATTTAAT GGTGAAGAAG ATAATAGTTA CTATTATCCA GTTATAAAAG 240  
 40 ACGGTAAAAT TGTTTATACT TTAACACTTA GTCCTAAAAA TAAAGATGAT TTAAACAAAT 300  
 CAAAAGAAGA CATGAATTAC AGTGTTAAAA TTTCAAACCTT CATCGCTAAA GATTTAGACC 360  
 AAATTAAAGA TAAAnATTCA AATATCACTG TTCTTACTGA TGraAAAGGG kTTTATTTTG 420  
 45 aAGAAGATGG cmaAGTTAGA TTAGTAAAAG CTACGCCTCT ACCTGGTAAT GTAAAAGAnA 480  
 AAGAAAGTGC TAAACAGTT TCAGCAAAAT TGaAACAAGA GTTaAAAAAT ACAGTAACAC 540  
 50 CTACTAAAGT TGAAGAAAAC GAaGCGATrC AAGAAGATCA AGTTCAATAT GAAAATACAT 600  
 TAAAAAaCTT CAAAATTwGA GaACAACAAT TCGATAACTC ATGGTGTGCA GGATTcAGTA 660  
 TGGCAGCATT ATTAAATGCa ACTAAAAATa CAGACACTTA TAATGCACAT GATATTATGC 720  
 55

	AAATGATTGA ATACGGTAAA TCACAAGGCA GAGATATTCA TTATCAAGAA GGCGTACCAT	840
	CATATGAACA AGTTGATCAA CTTACAAAAG ATAATGTAGG AATTATGATC CTTGCACmAA	900
5	GTGTATCTCA AAACCCTAAT GACCCACATT TAGGACATGC GCTAGCAGTT GTTGGTAATG	960
	CTAAAATTAA TGACCAAGAA AAACCTATTT ACTGGAATCC TTGGGATACA GAaTTATCAA	1020
	TCCAAGATGC AGATTCAAGC CTATTACATT TATCATTCOA TCGTGATTAT AACTGGTATG	1080
10	GTTCAATGAT AGGTTACKAA AAAGTAATAT AGATATTGAT TAAAGGCAGG TAAAACTATG	1140
	TATCAACTAC AATTTATAAA TTTAGTTTAC GACACAACCA AACTCACACA TCTAGAACAA	1200
15	ACCAATATCA ATTTATTCAT TGGTAATTGG AGTAATCATC AATTACAAAA ATCAATTTGT	1260
	ATACGTCATG GCGATGATAC AAGTCACAAT CAATATCATA TTCTTTTTAT AGATACGGCA	1320
	CATCAACGCA TTAAATTTTC ATCTATTGAT AATGAAGAAA TCATTTATAT TCTTGATTAT	1380
20	GATGATACAC AGCATATCCT CATGCAAACG TCATCCAAAC AAGGTATTGG CACTTCGCGA	1440
	CCAATCGTTT ATGAGCGCTT AGTATAACTA ATTTAAATGA TTTCACCTCA TAAAGCGGGT	1500
	TGGCGAGAAT TCAATTTCTC ACCAGCTCGT TTTTTCATTG TAATAATAAT CTTTAACATT	1560
25	TATTCTTTCT CTATTAATTT TTCTCAAAC ATCTTATCTT TATGATAATT AATTAAAATG	1620
	CCCTTTTAAA TTCTTATAAA ATAAAAAAGC CACCTATCGT CGCTAATAAA CGACGCAAGT	1680
30	GACTTAATAT CATATTCAA ATAACCTATG GGAATTTAGG GAATTGATCG AAGTCAGGAT	1740
	CACGTTTTTC TTAAACGCA TCACGGCCTT CTTTCGCTTC ATCAGTTGTG TAATAAAGCA	1800
	ATGTTGCATC CCCAGCCATT TGTGTAAAC CAGCTAAACC ATCTGTGTCA GCATTCATAG	1860
35	CTGCTTTAAG GAATCGTAAC GCTGTTGGTG AGTGTTCAT AATCTCTTTA CACCATTGCA	1920
	CAGTTTCATC TTCAACTTTC TCTAAAGGTA CCACTGTATT TACTAGACCC ATATCTAAAG	1980
	CTTCTGTGTC ATTGTATTGA CGACATAAGT ACCAAATTC ACGTGCTTTC TTATGTCCAA	2040
40	CGATACGTGC TAAATATCCT GAACCATAAC CCGCATCAAA TGAACCTACT TTAGGACCAG	2100
	TTTGTCCAAA AATAGCATTA TCAGCAGCAA TCGTTAAGTC ACAAACAACA TTTAGTACAT	2160
45	TACCGCCACC TACAGCATAA CCTTTTACCA TCGCGATAAC CGGTTTTGGA ATAATACGAA	2220
	TTAAACGCTG TAAATCTAAT ACATTTAAGC GAGGGATTG GTCTTCACCT ACATAACCAC	2280
	CATGTCCACG TTTCTTCTGG TCACCACCAG AACAGAATGC TAAATCACCT TCACCAGTTA	2340
50	ATACGATAAC TGAAACGTTT TGATCATCAC GTGCACGTGA AAATGCGTCA ATCATTTTCAG	2400
	CAACTGTTTT AGGTGTAAAC GCATTGCGTA CTTCAGGGCG ATTTATTGTT ACCTTAGCAA	2460
55	TCCCTTCGTA AAATTCATAT TTGATTTTAT CATATTCTCT AAGTGTTTCC CATTGTCTGT	2520



TCCACATGAA TTGTATGACC TGTAGCAGAA ATTAATTTAC ATTTACTATT AGGAATTAAA 2640  
 TTTGCCATTT TTTTCGCAAT CTGTACAAAT TTTTCATCAT ATTCTCCAGC TAATATTAAT 2700  
 5 GTTGGTACTT TAATTTCTnT CAGCGCGGCC ATAAGTTTGG CATTTG 2746

## (2) INFORMATION FOR SEQ ID NO: 327:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

TAATGTTTAG TTTATTAACA GTAAGTTCGT ATATCAATGT TTAGTGCTCC CAAAATTGA 60  
 20 AGTTTGAATT TTAAAAGCAT CTTGTAGAAT TTAGTTGTAT TTTTTTCAA GAAATTCATT 120  
 TTGATTATTT TTGATAATGA GCATTTTAAT AGTAATACAT GTTTATAGTG TGTAGTATAT 180  
 GTCTATACTA GTAGTAACTA TATAGAGAAA GTAGGAATAA ACTATGTCAC AAGATGTAAA 240  
 25 TGAATTAAGT AAGCAACCAA CGCCAGATAA AGCAGAAGAT AACGCATTTT TCCCATCACC 300  
 ATATTCCCTT AGTCAATATA CAGCACCTAA AACAGATTTT GATGGTGTG AACACAAAGG 360  
 TGCCTATAAA GATGGTAAAT GGAAAGTATT GATGATTGCT GCTGAAGAGC GATATGTATT 420  
 30 ATTGGAAAAT GGAAAAATGT TCTCTACGGG TAATCATCCT GTTGAAATGT TATTACCTTT 480  
 ACATCATTTA ATGGAAGCAG GTTTTGACGT TGATGTTGCG ACATTATCTG GTTATCCAGT 540  
 TAAATTAGAA TTATGGGCTA TGCCAACTGA AGACGAGGCA GTTATAAGTA CTTATAATAA 600  
 ATTGAAAGAA AAATTAAAC AGCCAAAAAA ATTAGCAGAT GTGATTAAAA ATGAATTAGG 660  
 ACCTGATTCA GACTATTTAT CTGTCTTTAT CCCAGGCGGA CATGCTGCAG TTGTTGGTAT 720  
 40 TTCTGAAAGT GAGGACGTTT AACAAACATT AGATTGGGCA TTAGACAATG ACCGCTTTAT 780  
 AGTTACATTA TGTCATGGAC CAGCAGCACT ACTTTCAGCA GGGCTTAACA GAGAAAAATC 840  
 TCCATTAGAA GGATACTCTG TTTGTGTCTt CCCTGaCTCA TTAGATGAAG GTGCAAATAT 900

## (2) INFORMATION FOR SEQ ID NO: 328:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3642 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

	ATCTGAnTnG AGATAGTGAT AATGTGTCAC CcAATTTTaa AcCtTTATTT TTaAGcGtTT	60
	CaTCaGCTAA CaCtTCaTTa TCtTTAGtCG cTTTATGcCC tTcTATTAAA CtCGGaACTa	120
5	AAAATGaTGa CtTTTCAACa CCGAACaCTA AAACaTTGtC tTTTTGATGG CCaTTAGACA	180
	CAATTTCCCC TGTTTGCTTC AAAGTAGCTT GCTTCTTGTA TTTATTTTCA ATATCTTTCT	240
	TGTTAAAAAC AGATTGTTGC ACAGTTTGAT TGGCATCTTT ATTTAGAACA ATGGCATCTG	300
10	CTTGCCACTT ATCAATGCCT TCTTTATTCA TATTGATAAG ACCATTCGCC AATCCAGATA	360
	ATAAAAATAG CAAGTAACTA ATCATCGTTA ACACACCAAT AATTAGTCCA AACTTCAATT	420
15	TGTTGCGCCG TATTTCAATC CAAGCTAAAA ACATGCATTT CTCTCCCTAC TACTATGATT	480
	TAAACATTGT TTATATTCTT AGATGCACGT ACGTCGTGTT GCGCTCTGTA ATGTTATACA	540
	TACACTTATC CTTCATTATA CCCGAACTTT TTATATTAAA AGCAAATTTA TGGAAAATGC	600
20	AATTAAATTGT CTATTATTTT TGTACGGTAC ATTTAAAAATT AAGGATCAAT TTAAAAACGC	660
	CTACATATAC CTTTAAGTAC ATGTAGACGT CCAATTCATA TATTATTTAA CTTCGCCTGT	720
	TTTAGGATCG AATTGCTTAA TAGCATTTTT ACGTAATTTA TCTTTTGCTT TrTCACCTGC	780
25	TTTATAGTTA TTGTTGTAAA TCGTAGCTTC CCAACTACCA TACATTGGGT TAGGGAAAAT	840
	GATATATTTT TTACCGAAAT CGTCTTTATG TTTTTCaATT AATGCTTCAC GAGATTCAGC	900
30	TGTAGCTTCT TTTGGATCTG TAAAGTCTAA TAAATTATCT CCAAATAGCA TGACAAGTTT	960
	ATGATCCTTT TGAACCATTT GTCTGCGTGA TTCTTTACTC TTATCATCTT TACCTTTTAG	1020
	TAAATATGA CTCTCTTAG CTTGAGGGAT ACCTTGTTGT TTTAAGTTCT TTTGTGTTGC	1080
35	CTTTAAATCT TTTTCTTTAT CTCTATCAGA AATATAGTAG ATATCGACAC CTTTTTTGTC	1140
	AGCATATTTT AAGAATTCTT TTGCGCCATA GACAGGTTTA GCTTTAGCAG cTTGTACCCA	1200
	TTCATGCCAA CCTTCTGGGA AAGGTTTATT ATGTATTGAT GCATAGCCTT GATATGGAGA	1260
40	ATTATCTAAA ACTGTTTCAT CTAAATCCAA AGCAATAGCT AACTTATGTT TACCTTTATT	1320
	CTTTTAAATC TCTTTATCTA ACTGTGTCTT TGCACTGTTA TAACCTTGTA AATATAATGC	1380
45	TTTTGCTTCA GCTGAATTTT GATACCAAGC CACTGCCATA ATATTTTGAT TACCAAGATT	1440
	CGCCTTTTGT GATGCTGGTA TAGAAGCTTG TTGCGTTTGT TGAACCTCAG CAGAACTTTT	1500
	GGCAAACGCT GTAGAATTTG TCGTTTGTGG TGCTGAAACT GTAACCGCTA CCGATAATGA	1560
50	TGCTATTGCA ATATACTTTG AAATTTTATT CATCTTATCA CCTCATGATT AATATTTAAA	1620
	ATACAGTTAA AATTATAAAT GCATTTATTT AATATTGCTA TACTATGAAA AGATATTTAA	1680
55	TATTATTTCT TGgAAAAGCT AACAAATATG TGAACATTTT ATATAAGCAT GATTTAATGG	1740

	GTGACTAGGC	CTTCCTATCA	GACATATTCA	CTCATCCACG	TATCATTATG	TGTACAGTGT	1860
	GCTATCTCTT	ATTTACCTAT	TGGAACAACC	ATAAACTCAT	CCATAGTTTA	CCTTTTATAA	1920
5	ATAGCAGTCC	TCACTCATAC	AATTTCTCAT	AAAAATCACA	ACGCTCCAAC	GTATTTCCAA	1980
	CTTACTTTCA	CCTATTTTAA	TTCATAAAAA	CGACACTTTA	ATTGTCATTA	TCCAATAATA	2040
	GCAAGACGTT	ATTATTGCAA	TCTTTTTTAT	AAAATAATAG	AATCATAGTA	TTGTCATTTA	2100
10	AAGATAAAGT	AAGAACGTTT	TTATTTTTCA	GATTTTTTTAA	ATTATTATGA	ATATCTAGTT	2160
	TTAGGAAGGA	AATTACATTG	AAAAACAAG	TTATTATTTT	GGGCCTCATG	TTATTTTCAC	2220
15	TATTTTTTGG	AGCCGGAAAT	TTAATATTCC	CGCCCATGCT	TGGCCATACA	GCGGGTCAAA	2280
	ATATGTGGAT	TGGTATGCTA	GGCTTTGCCC	TTACAGGCAT	ATnACTCCCC	TTTATTACTG	2340
	TTATTGTTGT	TGCATTTTAT	GATGAAGGTG	TTGAAAGTGT	AGGCAATCGT	ATACATCCAT	2400
20	GGTTCGGGTT	TATTTTTGCT	GTCGTGATTT	ACATGTCTAT	CGGAGCATT	TACGGTATTC	2460
	CACGTGCTGC	AAATGTCGCG	TACGAAATTG	GTACAAGACA	CATTTTACCT	GTGCATAACC	2520
	AATGGACTTT	AATTATATTC	GCAGCAATCT	TTTTTGCCAT	CGTTTACTGG	ATTAGTTTAA	2580
25	ATCCATCGAA	AATCGTTGAT	AATTTAGGTA	AATTATTAAC	ACCGTTATTA	CTATTAATGG	2640
	TCGCTCTATT	AAGTATTGCT	GTCATTTTCA	ACCCTGAATC	TGCACTAAGT	GCACCTAAGG	2700
30	ATAAATATAT	AACACATCCT	TTCATTTTCAG	GAAGTTTGGA	AGGCTATTTT	ACAATGGATC	2760
	TTGTTGCTGC	GTTAGCTTTT	TCCGTAGTCA	TTGTCAATGG	CTATAAGTTT	AAAGGCCTCA	2820
	CAGATCGCAT	GAAAATTTTA	AAATATGTCT	GCTTTTTCAGG	TCTTATTGCA	GCCATATTAC	2880
35	TTGGAATGAT	TTACTTTGCA	CTTGCATACG	TTGGGGCATC	AACAGCTCCA	GGAAACTTTA	2940
	AAGATGGTAC	AGATATATTG	ACGTACAAC	CATTACGATT	ATTTGGTTTC	TTCCGGTAACC	3000
	TCGTATTTGG	AATGACGGTT	ATCCTTGCA	GCCTAACAAC	ATGTATAGGA	CTCGTCAATG	3060
40	tTGCGCCACA	TTTACTAAGA	AACACGTACC	TAAGTTTTCT	TATAAAATAT	TCGCACTTAT	3120
	TTTctCTATC	ATAGGGTTCT	TATTTACAAC	ACTTGTTTAA	GAAATGATTT	TAAAAATTGC	3180
	TGTCCCATTA	TTGACTTTAA	TATATCCCGT	GTCGATTGCA	CTGTACTCA	TATCATTTGC	3240
45	TAACATGTTT	AGCACATTCA	GATTCAGTTG	GGCCTATCGA	CTCGCAACTG	TTATTACATT	3300
	GATTATTTCA	ATTTTACAAA	TACTAAATAG	TTCAACTTA	TTACACGGTG	TTATTTTGAA	3360
50	ATCGTTTATG	ATGTTACCTT	TAGCAGATAT	CGATTTAGCT	TGGCTTGAC	CATTCATGCT	3420
	CCTTGCTATT	ATCGGTTTCA	TAATCGATGT	ATTTATACGC	CGTCCGAAAC	AAGCGACAAC	3480
55	TTAATAAATG	CTCACTGCCT	AGTAATGATT	GACCCATCGT	TACTAgGCTT	TTtctATATGA	3540

TAAAAAATCC TAGCTGTTAT TCAAAAATAT TAGTTTTTAA AA

3642

(2) INFORMATION FOR SEQ ID NO: 329:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2187 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

15	TTATTGATAT TGAAAATTCA AAAACTGCAA CACATCTTAT CACTAAAAAC CAACACTATC	60
	ATCAAACTGA CACTCATTTT GAACAGTATA AGAAATTTAT TTTAGATTCA GGTATATCAT	120
	CAACACAATT TGTATATAAT AACCTGTCTG TAAGCGGATT TAAATATACT AATGATGGTA	180
20	AGAATCCAAT TCAATTATCT GACATAGTGT ATCACTTAAT CGCATTATTA CGATATGGCG	240
	GTGGCATTAG CTATCAATTA TTAGATGACC ATTCAAATTA TATTTCTTG TACAACAAAT	300
	ATGGTAGCCC CCTACCATTA ATGCATTTAT ATAAATGTT TAGACCTTTT GTTAATGAAG	360
25	ATATTGAAAT TACAAATAAT TATGTATTGA GTCGTAAAGA TAATAACTAC CATTTCTTAT	420
	TATTCAATAA AATTAATGAT CGATATATGT CAGACGTAAA ACAAGATTTC ATTTTCCATA	480
30	ATGAATTACC TCAAGACTCT TTGATGATTA TTAAACATT GAATCATGAA CATGGTTCAA	540
	TTCAACATTT GCTTCCAATA AGCGATCAAC TTGTTTATAT AGAAAAAGAA ATTTTAGATG	600
	AATTAGACAA AACGAATTAC CCTAAAACGG AGCTTGCAGT TCAAGAAGAA ACTGGTCGTA	660
35	CATTTGAAGT CAAGTTAAAT CACGACGAGG TTAAATATAT TTGCTTTAAA CCAAGCTAAA	720
	TACTAACAGT CCTCTGTGT TTAGTTTCTT ACGTTAAAGG CTATTTATAT CATAAGGAGA	780
	TGATTTGTAA TGAGTAACTC ACAAGCAATT CAAGCAATTG AAAACGTGTT AGTAACGTCA	840
40	AAAGTTGGTG TATTATCAAC TGCATATAAT AATAAaCCTA ATAGTAGATA TATGGTCTTT	900
	TATAATGATG GTCTTACTTT ATATACTAAA ACGAATATCC ATTCTGCTAA GGTCAAAGAA	960
45	ATTAAAGATA ATCCAGCAGC ATATGTTTTG TTAGGCTATA ACGACACAAC TAATCGTAGT	1020
	TTCGTTGAGA TGGAAGCGAC AATCGAAATC GTTACAGAAC AAGAAGTGAT TGATTGGCTA	1080
	TGGGAAACAC AAGACAAAAG CTTTTTCAGT TCAAAAGAAG ACCCAGAGTT ATGTGTTTTA	1140
50	AGAGTAGTTC CGCAATCCAT TAAGCTAATG AATGATAAAT CATTAGATAC ACCTATCAAA	1200
	ATCGATTTAT AACACAAAGT GTATATAGGA AATAACTTTT ATGAATTCTA GATATAACAA	1260
55	TGTTAAATAC TTAAATAAC TCGCTATAAT TAAAGTGTTT AATATGTTTA CAATTCAATT	1320

TTTAAATTAA TTTTATGTAA TATAAATACT GCATTGCAA ACTGTTGCAC TTTTAGGTAT 1440  
AACAGAATTA ACTACATTTA AGGAGATTGA TGAACCATGA AAAAGAAAAA AGGTTTTGGT 1500  
5 CTTGGTATTA GTTTAATCGC CATCATGTTA ATTGTATGTA TTGTATTAGT AATCATGATG 1560  
ATGACTGGCG GAAAGAAAGA TACATACTAT GGAATTATGA AAGATAATAC TACTATTGAA 1620  
AAAAATGATTA GTGAAAAAGA TGAAAGTATT GAAAAAATG TTAAATTACC TTCAGATTCA 1680  
10 GATGTTAAAG TTAAAAAAGG TGATTTTGTA ATTGTTTATA AATTAGCAGA TTCAGATAAA 1740  
ATTGTTAAAG TTAAAAAAGT TGACCATGAC GATGTACCAC ATGGTTTAAT GATGAAAATT 1800  
15 CATGACATGG GCAAATGCA CATGAAACAC TAATTGTAAT TTAAATTACA AATTTTAGTT 1860  
GCCATCAAGG TATATACGAG TAAAGCAGC GGTAAGTTGA TTTCCAATTT GGAATCATTT 1920  
TACTGCTGCT TTTTATATTT GAAATACTTT CATATTGAAT AGCTCCACTT GCCGTTGCGC 1980  
20 TgcgCTTTGC GCATGCATAA AAGCCCCTAA CAACCTGAGG TCACTGCGCT CCGGTTGCGC 2040  
TGcgCTTTAG CGCATGCATA AAAGCCCCTA ACAACCTGAG GTCACTACGC TTCGGTTCGC 2100  
CTGCGCTTTA GCGCATGCAT AAAAGTCCCT AACAACCTGA GGTCACCTACG CTTCCGTTTCG 2160  
25 CCTGCGCTTT AnCATGGCCA TAAAAGC 2187

(2) INFORMATION FOR SEQ ID NO: 330:

(i) SEQUENCE CHARACTERISTICS:  
30 (A) LENGTH: 1788 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

CCnCCTTTTT AAACCTGGnG AAATGtmCAG tTTTGATGGt ATTGGGtTCT TTAGTATTAT 60  
40 GCTTAAGATA GAGTGTAATA CGCTCCTGTT GTTCTTTATA GTATATTGCT TTTTGTTTTT 120  
CTTTTTTCGTC ATATTTCACT TTAAATAAAA TGA CTGATGC AACTATACAT ATACATAAAA 180  
TGACACCAAT AATTATAAAA ACATGTTTTT TGTTTCATTAA AGTAACTCCT AAAATGTGGT 240  
45 GGAAAATGAA AAAATTTTTA TAATCTATAA TTATGAACAT TACAGATTAT AAACCATAAC 300  
ACTAACATCG TCGCCTTCAT TAACTTGATT GTTAAATCA GCAATTACTG AGAATTGTAC 360  
50 AAGGAATAAT TGCCTATTAT GCCCTCATGT AATTATTGCC TTA CTAACAT TAACAAAATC 420  
GATAGCTATT ACATTAAATG CCTATACCCC AGACCTCAA CACCTTTTTA TACAGGACGC 480  
ACTGTAATTT CATyyACGTT AACGTGCTTT GGTGTGTGTA ATGCATATAA TACTGCTTCT 540

55

GTATCTACCA TTCCTGGAGA AATGCTTGTT ACTTTAACGC CTGTCTTTGC CAACTCTTTT 660  
 TCTAATCCTT GAGTAATAGT GTGAACTGCT GCTTTCGTCG CACTATAAAT CGTACTACTT 720  
 5 TTCGTTACTT CAAAGCCAGA AATAGATGcA ATGTTAATAA GATGGCCACT TGATTGTTCT 780  
 AACATAGTTG GTAATGCAGC CTGTGCCGTA TATAAAGTGC CTTTGATATT CACATCAATC 840  
 10 ATACTATCCC ACTCATCTAC TTGATAATCA GTAATCTTAG ACGACAACAT TTGCCCCGCA 900  
 CTATTGATAA CAATATCCAA ACCACCGAAT GTTTGTTGTG CAATTTTTTAT CAATTCATCG 960  
 ACTTCTTCTT TATTCGTTAC ATCTGTTGGC ACTACCTTCA CACTATCTTG TGACAATTGA 1020  
 15 TTCGCTACGT TTTGTAATTT TTCTTTATTT CTACCTGCTA AGACAACTTT TGCCCCCTCT 1080  
 TCATGTAGTA ATGTTGCAAT TGCTTCTCCA ATACCACTAC CTGCACCTGT AACTACTGCT 1140  
 ACTTTATCTG TTAATACTGT CATAATGATC GACTCCTTTG ATTCTTTTTTA TTTTTTCAGG 1200  
 20 GTAAATCATA AATACATATT ACTTTTAAAA AGCGTATCAC AATTCATATA ACGGTCATAA 1260  
 TAACTCGCTT CATTTTCATA GATAAATTAC ATTACAAGCC ATTCGAAACA TACAATTAAT 1320  
 CGTTGCTTAT ATTTTTTATT TTTAAAAATG TTGAAAAATC GTCACCTCTT TATTGTAAAA 1380  
 25 ACATTATATT AGTAATAAAG TTAATACTGT GnATTTaTCA TTCGATTGAA TGATTAGAGG 1440  
 GAGGAATAAA ACgTGACATA TCATGAGCGT GTTTTAGCAT TAAGAGCAGA AAGTAAAAGA 1500  
 30 ACCGCAITTG ATTTTCGATT CGAAGATTTA TTTAGCAAAG AAGAATGGcT AAGTATGTCT 1560  
 CTTGCAGAAA GACAAAAAGC TGAAAAAGCA TTTGACACG AgTTAAAAAT ATGGACGATG 1620  
 TAAGAATGCC CTTCTCAAGT GTCCATGACG CCCAAGTAAA ATTATATAAT GTTGTATATT 1680  
 35 CTTATAACGG CaTTAAACGT AATTTTAAAC AAGTTGAAAA TGGAAGGATT CTAATATCAT 1740  
 TTCGTTTATA TATnGCAGAC CATGGATAGA ATTTTnTATG GTnAATCC 1788

## (2) INFORMATION FOR SEQ ID NO: 331:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1341 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

50 TAAGCCAAAT CCATTGTTAC CAGAAGAAGT TCGCTTCATC TTAAAAACAA TGGGTCTTAA 60  
 AGAAAAGACT ATCGATGTTG AACTCGAAGT TGGCGAGCAA GTTCGTATTA AATCAGGTCC 120  
 ATTTGCGAAT CAAGTTGGTG AAGTTCAAGA AATTGAAACA GaTAAGTTTA AGCTAACAGT 180  
 55

GCTTTAATTA ACAATTAAAG TTATTAACT AACCAAAAGA TAAAAAAGAG TATTGATTTT 300  
 TTAATTAGAA AAGTGTTAAA ATTATGTGGT cGcGCTTTTA GAGCGCCCAT TTCGTcACGA 360  
 5 AATGTTAAGA GTGGGAGGGC AAAACTGAGC CCTGTGACCA CATCACGATA TCAAGGAGGT 420  
 GCACATCGTG GCTAAAAAAG TAGATAAAGT TGTTAAATTA CAAATTCCTG CAGGTAAAGC 480  
 GAATCCAGCA CCACCAGTTG GTCCAGCATT AGGTCAAGCA GGTGTGAACA TCATGGGATT 540  
 10 CTGTAAAGAG TTCAATGCAC GACTCAAGA TCAAGCAGGT TTAATTATTC CGGTAGAAAT 600  
 CAGTGTTTAT GAAGATCGTT CATTTACATT TATTACaAAA ACTCCACCGG CTCCaGTATT 660  
 15 ACTTAAAAAA GCAGCTGGTA TTGAAAAAGG TTCAGGCGAA CCAAACAAAA CTAAAGTTGC 720  
 TACAGTAACT AAAGATCAAG TACGCGAAAT TGCTAACAGC AAAATGCAAG ACTTAAACGC 780  
 TGCTGACGAA GAAGCAGCTA TCGGTATTAT CGAAGGTACT GCACGTAGTA TGGGTATCGT 840  
 20 TGTAGAATAA TTTTACGAAT ATTAAATTTG ATTACATGAT TTAAACGATG AAGCAGATAA 900  
 CAGAGATAAT AATGATGAAT TATAAATATA ATCTGAATGA CTAGATTAAT GATTGATTTA 960  
 TTCATAAGAT TAATTCTTCT GTTGTCTGcy CTTAACTTGC ATATAGCAAG TAATGTGGGA 1020  
 25 GGAAATTCCG CTAACCAC TAAAGGAGGA ACTATAAATG GCTAAAAAAG GTAAAAAGTA 1080  
 TCAAGAAGCA GCTAGTAAAG TTGACCGTAC TCAGCACTAC AGTGTGAAG AAGCAATTAA 1140  
 30 ATTAGCTAAA GAAACAAGCA TTGCTAACTT TGACGCTTCT GTTGAAGTTG CAITCCGTTT 1200  
 AGGAATTGAT ACACGTAAAA ATGACCAACA AATCCGTGGT GCAGTTGTAT TACCAAACGG 1260  
 AACTGGTAAA TCACAAAGTG TATTAGTATT CGCTAAAGGT GACAAAATTG CTGAAGCTGA 1320  
 35 AGCAGCAGGT GCTGACTATG T 1341

(2) INFORMATION FOR SEQ ID NO: 332:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5136 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 332:

CTCTAAATCT TCAATAGGTA ACTTCATTTT AATAATTCCC ATTTGAACAA TTGCTTCTTC 60  
 50 ATCATTGACA ATTTCTAAAA CTTCAACCCTT TTGACCGTAA GAKAATACTT TkACTTCATC 120  
 ACCAGCAACA ATTTTATCGT ATTTTGTCTT TTGTACATTT TGCTTTATAG ATTTGCTTC 180  
 ATAATGATCA TCTAATCGTT TCTTCTTATC AATCAATTCA TGTTCTTTAA CATCTGCACC 240  
 55

	CTTAATCTTC TGATTTGCTT TKTCTTTTCGC TTCCTCTATT AGAGACTTTT CATAATTTTG	360
	GAATTGTTGA TACTGCTTAG ATAAATCATC ATGCACTTgt CCGCKtskTT TACAAGACGG	420
5	TCCAGkTCTA ACCTCTGTGT CTCTACACGT TTGTAATTAC GCTCTAATGA TTCAATCATT	480
	TCATTTATTT CTTTTTCATC AGTACCAATC ATCGTCTTAG CTTATTAAT AATATTCAAA	540
	CTAAGACCTA ACTTTTTAGA AATGTCAAAA GCATTTGAAC GACCCGGCAC ACCCATTAAAT	600
10	AACTTATACG TTGGACTCAA AGTATCTACA TCAAATTCTA CACTCGCATT CATAACGCCT	660
	TCTCGATTAT AACTATATGC TTTAAGTTCA GGATAGTGCg TCGTTGCCAT TACTAGAGAA	720
15	CCAATTTTTC TAACATGATC TAAAATGCTC ATTGCTAATG CAGCACCTTC ACTCGGATCT	780
	GTACCTGCAC CTAATTCATC AAATAAACT AACTATGTT TGTCTGCATG CTTTAAAT	840
	TCAACTATAT TCGTCATATG AGATGAAAAA GTTGATAATG ATTGTTCTAT TGATTGTTCA	900
20	TCTCCGATAT CGCAATATAC ATTTTTAAAT AACTTAACT GACTACCATC AAGTGTGGGA	960
	ATCAACAATC CTGATTGAGC CATAACAATA ATTAAACCTA ATGTTTTTAA TGTTACAGTT	1020
	TTACCACCTG TATTCGGTCC TGTAATAATT ACCGTTTCAA TATCTTCCAT AAATTCGATG	1080
25	GTATTAGCTA CAACAGTCTC ACGATTTAAT AATGGATGGT ATGCTTTAGG TAAATATACA	1140
	GTACGGTCCT CTTTAAATAT CGGCTTTGTT CCTTTAATAC TTCTACTATA TCTCGCTTTT	1200
30	GCGATTAAAA AATCTAACTG ACCCATGACT TGTCTGCCA CAAGTAGTGC ATCTTTGTCC	1260
	GCAGCCACAT AACCAGTTAG TTGCGTTAAA ATGCGTTCTT TTTCAATTGC TTCGTCAATGA	1320
	CGTAATCGAC TAATTTGATT ATTCATTTCA ACAACTGATG ATGGCTCAAT ATACAATGTT	1380
35	TGTCCTGAAG CAGATTGATC ATGTACAATC CCATTAAAAT CTTGTCGATA TTCAGCTTTG	1440
	ACAGGTATAA CGTTTCTTTC ATTCCTAACT GTTACAATAG CATCTGATAA TTTTCTGA	1500
	TTTGCTTGGC TTTTAAACAT ACGGTCCAAA TTTTGTCTAA TACGTTGATT CGTGCTAGAA	1560
40	ATTTTACTTC TAATCCCTTG CAATTCATAA CTCGCATTAT CATATAAATC ATACGTATCG	1620
	CATGTTTCAT TTATTTGTTG AAAAAGATCA GTTAACACAG GTAATTGATT CATCTGTCA	1680
45	TCTAATATTG GGTATTTAAC ACCTTCATCT TCTTCAACCA ATTGATTATA AAATGTCTTG	1740
	AATTGATTTT GTACTTGAAT TAATCTTTTT ATCAAGTTAA GCTCTGATAC ATTTAAAACG	1800
	CCGCCAATAT CAGCGCGATG AATGAATGCT GATACTTTAG ATAAGCCACT CAAGCTTGGT	1860
50	AAACGATGCT TATTATAGAT TTGAGCAATC TCATCCGTTT CTTCATTG AAAAACAACC	1920
	GTTTCAAAAT TAGTAGCTGG CATCATTTGA TTGACCTTTT CCAAGCCTAA GTCACTAATA	1980
55	GTTTCATTGG CAACGAGTGA TTTTATTTTT TCAAATTCTA AGACGTCTAA TGTTTTTTGT	2040



	TCACGCGATA ATGCGTTAAT CACTCTATCT TTTGTtACAA ATCCTTTTtG CGCAGTTGtA	2160
5	CGCCATAATT CATAAAATCT AAATGATTtG TATGATGCGC ATCAGTGTtA ATAGTTAATT	2220
	TCACATTtGG ATATTTACGA ACGATATCAG CGCTCAGATC CAGTCGATGT GGATTGGCAT	2280
	TAATTTCTAA TACTGTATTtC GTTTCTTCAG CTAATGCCAT TAATTGTtCA ATATTtCGGT	2340
10	TATAACCATC TCTTCTACCT ATAATACGCC CTGTTGGATG CGCTATATGT CGCACGTATG	2400
	GATTGCGACA TGCATTAGCT AATCGTTCCA TAATTTGTtC TTCTGATTGG TTAAAGCTTT	2460
	GATGAATAGC TCCAATTACA TAATCAAGTT GTGCTAAAtT TTCATCATCA TAATCCAGCG	2520
15	AGCCATCAGG TAATATATCC ATTTCTGTAC CTGAATAAAtT ATCAATTtCA CTATATTCTT	2580
	TATCTAAAGC CTTAATTTCT TCGTTTTGTtC TTAAAGTCTT TTCTACTTGT AAGCCATTAG	2640
	CAACACGTAA ACTTTGTGAA TGATCAGTAA TTACCATGAA TTTATAACCT TTTGCGATAT	2700
20	TTGCTTCTAC CATGTCTCGA ATAGAAAACG CACCATCACT ATACGTTGTA TGCATATGAA	2760
	TATCACCATT AATATCATCT ATTGTAATGA TATTACTTAG ATCTTTATCA AATTtCGCTAC	2820
25	CATCTTCTCG CATAGCAGGT GGTATAAAAtT TCACATTAAAtT ATGTTtCATAT ATCTTGGCTT	2880
	CACTATCATA TTGAATTAAT GTACCATCAG CTTGTTCAAT TCCATATTCA CTTACTTTTT	2940
	CATCACGTGC TTTAGCAAGT TGTCGAATTtC TTATATTATG TTCTTTTGAC CCAGTAAAtT	3000
30	GCTGCAATGT ATGATAAAAA GCACTTGGTT CAATTAATCG AAAATCGACA CCAATCGTTT	3060
	CATCATCATA CGCTAATTCT AATGAAACTT TTGTGTTCCC CACTGCAACT TCTTTTACTT	3120
	TATTGGGAAT ATTTAATAAT TGCTGCTGCA CTGCTTTTGG GTTATCGGTA CTTATTATGA	3180
35	AATCTAAATC TTTGCTCATT TCTTTAAAC GACGGAAGCT TCCTGCAGAT GAATATTGAT	3240
	CGATATAATT TAATGTATCT ATATAATCAA TGATTtCTTG ATTAAGTCTT CTCATTtGAT	3300
	CAATTGGATA TCTATCTTtC TTAGCACCAA GTTGTtTCAC AGCTTCTAAT ATGTTTtGTT	3360
40	CCGTTTTCTT AGCAAATCCG CTTAATTtCAC TAACTTTtCC ATTTTCACAA GCAACTTGAA	3420
	GTGACGCTTT ATCAACAATA TTCAACTCTT TATATAGCTT AGCAATTTtC TTGCTTCCAAtT	3480
45	GTCCTTGAAT TTTCAAAAGT GGAATAAGAC CTTCCGGAAC TTCTtCCTGT AATTGCTGTA	3540
	AATACTGAGA TTCACCGGTC TCACGGTAAT CATTGATTAC TTCTGCAACA CCTTTACCAA	3600
	TGCCTTTtAA CTCCGTtACA TCAGATATTT CATCTAATGG TCGTTtCATCT AATTCAAGAC	3660
50	TTTGAGCTGC TTTTCGaTAC GCTGaTATTT TAAAAGTATT TTCCCCTTTT AATTCCATAT	3720
	AAGTAGCAAT TTGTtCTAAT AGTTTGATAA CATCTTTTTT TGTCATAATA ACACTCCATA	3780
55	AAAAGAAGAC CAGGACGTAT CATTAAATATA TACCTTTGTtC CTGACCTCTT ATGATAATTT	3840

TAGATATTTT AAGCTGATGT TGTAATGCTT CGTTAGGATA TAATGCCAAT AGATATAACG 3960  
 TAAAGTGTA GACAATTATC GTCATAAACA CACCAACTAT CATTCCCATT GCACGACTGA 4020  
 5 AAATATGAAT GTTTGTATAC GCTATTATTT TATCAAAAGT TACGATAATT AGATATAAAA 4080  
 TGAACCTACA AAACAATGTA ATCATTAAAA AAGCTACAAT CGCTTCAAAT CGATTTTGTA 4140  
 10 GATGATTAAA ATGAAACGCA AAAGTTGTAT TAAATGCTGT TGTTTTAGGA TATGGAATAA 4200  
 ATACAATTAA TCTTCTTACA ATAGATTGTG AAAATTGACT GGCAATCCAC AATGATACAA 4260  
 nCGTTGCACT CAAATGTATC ATAGATAACC AAAAACCTCG TCTGAATCCA ACGATGACAA 4320  
 15 AATACACAAA GAAATGATT ATGATAAAAT CAATGACCAT TTATTGCTCA CGCTGCTGCA 4380  
 ATTTGTGAAT TTGTTGTTTC AAACGTCGAT TTTTTTCTT CTAGTAGTAC TTTTTCATGC 4440  
 ATAATATTCA CAGCAGTTAG TATTGCTTTT CTGAAGTAT CTAAACCTGC TGCTTTATAC 4500  
 20 CCTAATCTT TTATTTTATC ATCAACTAAA TGTGCTACAT ATCGTATGTG CTCTGGGTTA 4560  
 TCTTCCCCAA CAATTGTAAA AAGCTGATCA TTAATTGATA CATTACCTT GTTTTTAAAC 4620  
 TGTGTCATTT ATAATTTCTC CTGATCCTTT TTTTAAAATC TAAATTCACG TTATAAAATA 4680  
 25 TGACTGGATA GTTTGTCTGA ATTTGATACT AATATTGTGA TATTGCAATT ATGATAAAAC 4740  
 AACAACACAA TCTCTATAGA TGACTTAATG TTCTTTTTAT AATGAAATAA TGTAAGAAT 4800  
 30 TTTCTATTCA ATACTTTATC ATGTTTAAAT TGTGTCAC TAACATTTTC ATAAACATTA 4860  
 TACATGACCA CTATGTATTT TGTAAGTATC CGCAATTAAT TCTTACAAC ATACATAAAT 4920  
 GTTCTGACG TTATTATCAT TTATGATATG ATTATTTTTG CTAAAGACAA TGAAATTTTA 4980  
 35 TGAAAGGATT TACACAAATG GCGAATATCG TTTTAAATTT GTCGGATAAA GACATAACGA 5040  
 CATTAATGTC ACGCATTTCT TTTGATACTG AGAATTTACC TCAAGGnATG AnAGCACGTG 5100  
 CAAAGTATCC AAAATACAAC TGTAATATT TACCAT 5136

40 (2) INFORMATION FOR SEQ ID NO: 333:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4239 base pairs  
 (B) TYPE: nucleic acid  
 45 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 333:

GGCCAAAATT GCACCTCCAC TTTCCTTTGA ACAAATCTAT TTAAACGCA CATTCCCATT 60  
 ATGACAAAATT AAATGTGAGT nACATTGTGTT TGTATTTTAA CATGACTACT AACGCAAACT 120

55

	GGTATCTTTC AAAGATAAAA TCTTAATAAT TTCTTAGTAA ACTCTTTTCT CTAGATTTAT	240
	CACAATATTA TATAGACCTA TTTTATTTTG ACGTAAGTTG CTAGTATCTT CAAACAAAAA	300
5	ACCTTTATAA AAATTCATAC CTTTATGCTA TCGCTGTAGG CTCATTAAC TGTACATAT	360
	AATTCTTAAC TATCCTTTGA TGATTGTTTT ATTAGATTGT TTCGTTGATG GATACTTTCA	420
	CGAATTTCTA TAGTTCAATG CTACTAAAAA AACAGCCCAA AACTTTAATT TGTTTTGGAC	480
10	TGTTTTATAA TTATGCTTGC GATGGTGTTC TAGTTTCTGA AGTTTGTTCA GCAATGTCAT	540
	ATTTAAACTC TTTACCATCA TGATCTACTG TAACTTTCTT ACCTTCAATT TGATTACCAT	600
15	CTAATATTAA TTCACTTAAA TTATCTTCGA TAGTTTTTTG TATCGCTCTA ATTAATGGTC	660
	TTGCACCATA TTCTGGATCA TATCCTTCTT CTGCGATTTT GTCTTTCGCT TTATCAGTTA	720
	CAATAATATT TATGTTTTGT TCAGATAATC GATTTGTTAA TTTATTAACC ATCATTGTTA	780
20	CAATTTCTTT TAATTCCTTCT TTGTTAGTT TATGGAATAC AATGATATCA TCTACACGGT	840
	TTAAAAATTC TGGACGGAAT GAATTTTTTA ATTCTTTTAA CATCGTTTTT CGAATTGTTT	900
	CATAATCTTG TCCATCACTT GAACCACCGA ATCCAGCAAA TCGTTGATCT TGTAATCTTT	960
25	GTGCCCCAAC GTTTGATGTC ATTATGATAA TTGTATTTCT GAAATCAACT GTACGTCCTT	1020
	TTGTATCTGT CAAATGTCCA TCATCTAAAA CTGTGAATAG AATATTAAAT ACATCTGGAT	1080
30	GAGCTTTTTT AATTTCATCA AATAAAATTA CAGAATATGG TTTACGTCTA ACTTTTTT CAG	1140
	TTAATTGTCC ACCATCATCA TGACCAACAT ATCCTGGAGG AGCACCAACT AATCGGCTCA	1200
	CTGCGTGTTT TTCCATAAAT TCACTCATGT CTACACGGAT CATCGCATCA TCATCGCCAA	1260
35	ACATTGATTC AGCTAAAGCT CTAGCTAATT CAGTTTTACC AACACCAGTT GGTCCAAGGA	1320
	AGATAAAGCT ACCAATTGGT CGTTTAGGAT CTTTTAACCC TGCACGGGCA CGTCTAACCG	1380
	CTTFACTGAT TGAATTAACA GCATCTTTTT GCCCAATAAC TCTCTCATGT AATGTATCTT	1440
40	CTAGACTAAG AAGTTTTTCA GATTCTGTTT CATTGATTTT AGTTAATGGG ATACCTGTCC	1500
	ATCCTGCAAT AACTTCAGCA ATATCTTCTT CTGACAAATGA AGTTGACATG CCATTTTGTG	1560
45	CATTCTTCCA TTCATTTTTA GCTTCTTCAT ATTGCTTTTC AAGTTTGTG TGTATATCAC	1620
	GCAGgTTAGC AGCATTTTCA AACTCTTGAG CATGTACTGC GGCATCTTTT TCATTTTTAA	1680
	CTTTTCAAT TTCTTGTTCA ATTTCTTTTA AATTATTAGG TGTCGTATGA CTCTTAAGTC	1740
50	TTACTTTAGA ACTTGCTTCA TCAATTAAAT CAATTGCTTT ATCTGGTAAG AAACGATCTG	1800
	AAACGTATCT GTTACTTAAT TTAACAGCTG CTTCAATAGC TTCGTCTGAA ATATTAATAC	1860
55	GATGGTGTGC TTCGTAACGA TCTCTTAATC CTTTTAAAT AGCAACTGTA TCTACTACTG	1920

	TTTTGCGATA	TTCATCTAAT	GTAGTAGCAC	CAATACATTG	TAATTCACCA	CGTGCTAATG	2040
	CCGGCTTCAA	AATATTCGAA	GCATCGATAG	CACCTTCAGC	ACCACCAGCA	CCAACTAAAG	2100
5	TATGCAACTC	ATCAATAAAT	AGGATGACAT	TACCTGCTTG	TTGGATTTCT	TCCATAACCT	2160
	TTTTCAGACG	CTCTTCAAAT	TCACCACGAT	ATTAGTAGTACC	TGCAACTACT	GTTCCCATAT	2220
	CTAAAGACAT	AACACGCTTA	TCTTTTAATG	TCTCTGGTAC	CTCATTATTCT	ACTATGGCTT	2280
10	GCGCTAAACC	TTCAGCAATA	GCAGTTTTAC	CAACACCTGG	CTCTCCAATA	AGCACAGGAT	2340
	TGTTTTTCGT	ACGTCTACTT	AATACTTCAA	TTACACGTGT	AATTTCTTTA	TCACGTCCTA	2400
15	TAACAGGATC	TAATGTACCG	TCTTTGGCAA	TGACTGTATA	GTCACGAGCT	AAACTATCTA	2460
	AAGTTGGAGT	ATTATTGAC	TTACTAGCTT	GTGCATTTTT	ATTACTCATT	TCAGGGTTTC	2520
	CTAAAGCTTT	CACAACTTGT	GCACGTGCTT	TAGTAATATT	TAAATCTAGA	TTTGCAAAAA	2580
20	CTCTTGCTGC	AACACCTTCA	TTTTACAGAA	TCAAGCCTAA	TAAAATATGT	TCCGTTCCAA	2640
	CAAAATTGTG	ATGTAATTTT	CTAGCTTCAT	CCATCGATAA	TTCAATGACT	TTTTTAGCTC	2700
	TAGGTGTATA	ATGCAATGTA	CCAACATGAT	CTTGACCATG	TCCGATTAAT	TTTTCAACTT	2760
25	CTTCAATTAC	TTTATCTTCA	GTGATATTAA	AACTTTCTAA	TACTTTTGCA	GCAATTCCTT	2820
	CAGGTTCTTT	CATTAACCCC	AATAATAGGT	GTTCTGTTCC	TATATTTGAA	TGATTTAAAC	2880
	GAATTGCTTC	TTCTTGGGCA	TGTGCTAATA	CAGCGCTGTC	ACGCTCAGTT	AATCTACCAA	2940
30	ATAACATAAA	TAATGACCTC	CTACTTTATA	TGTTCTCTTA	GTATATCTGC	TCGTTTTTCT	3000
	TTTACAGATT	TGTCATCTTC	TTTATCTAAT	AAAAATGGTG	ACTGTATAGC	TACCATCAAT	3060
35	TCATTAAATT	TAAAGTTTTG	TAATTCAATG	TAATTTAAAT	CTATACCAAG	TTTAACTCGC	3120
	TTAATCTATA	AGAAGCCTCT	TCCATAGTTA	TCATTCTACA	GTTTTGTAAA	ATACCTAGCG	3180
	AGCGAAAAAC	ACGGTCTTGT	GTTTCTAATT	GATTATAAGT	GTCTAACTTT	TGTCGTATTT	3240
40	GTTTTTCTTC	ATGAATGATT	TGATTAACAA	CTTCTGTATA	TGTTTCTATG	ATTTCTAACT	3300
	CAGATTTACC	AAGTGTAAGT	TGGTTGGATA	CTTGATAAGT	ATGTCCATAA	ACTTGCGAAC	3360
	CTTCACCGTA	AATACCTCTG	ATTGTATATC	CAAAACGATT	AATGGTTTGA	GCAATCCGTG	3420
45	TCATTCTTTT	CATAATAGAT	AGACCTGGCA	AATGTAGCAT	CACGCTTGCT	CTCATACCAG	3480
	TACCTATATT	GGTAGGACAT	GTAGTTAAAT	AACCAAGTTG	TTCATCATAA	CTTATATCAA	3540
50	GGCTTCGATC	TAATTCATCA	TCAATTGATG	AAGCTTGATT	ATATAAAGCC	TGTAATGTCTG	3600
	TGTCAGTTCC	CATAGCTTGA	ATACGAATAT	GGTCCTCTTC	ATTTATCATG	ACACTTAAAG	3660
	ATTCATCGTC	ATTCACTAAT	ACTGCGGCTG	CTGGTTGTTT	TATTAGTTCA	GGACTAATCA	3720
55							

CAAAGTTTGG CAAGGCATCT TGTACCTCAT TTATAACTCT AAATCCcATC ATTTTCAGTA 3840  
 GCATACATTA GTGGATGCAC ATGATTTTCT AAAcTACGCG CTAACCGAAT TCTAGAAGAC 3900  
 5 ATAACAATTG GTGTTTCTTC ATTACTTTTC ATCCATTGGC TGATATTATC ATGAATATTA 3960  
 TCGTCATCA TGTTGCACCT CACTCTCAGC TTTTAGTGCT TTAATTTTCAT CTCTAACAAT 4020  
 GGCTGCTTCC TCAAAATCTT GGATTTCaAT AAGTTTTTTC AAATATTTCAT TCTTTTCTTC 4080  
 10 GATTTTTCGC TTAAAGCTA TCTTTTATG TGAAGAATGT GGTGTCTTTC CAACGTGCTC 4140  
 AAATTGTCCA CCTTGAACTC TGCGGACGAT ATCAATGATG TCATCTTTAA ATGTTGcATA 4200  
 15 ACAATTAGCA CACCCAAATT TACCAACATG TGCAATATC 4239

(2) INFORMATION FOR SEQ ID NO: 334:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1245 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 334:

CAAAACTTTG ATTATGTCGT GTCAGTTATT AATGCATACA GTAATGCATC GTTGTTATTT 60  
 TTTATTTATT TACTATTTAT TATGTACCAA ATCAGCCTTc CAGTAAATA GGTCAAATTA 120  
 30 CTGATTTTCT AAATATAAAA TGCCTCCTAA TAACATACTA TTAGTACATC ATTAAGAGGC 180  
 TCTTGTTGTA TTTGCATACT AAGCGCTCAA ATTTAAATTT AAGATGAAGA TTCTTGCAAT 240  
 35 AATATTTCTA TATTTGTGCA TATATCCAAT GGATCTTCCA TTGGCAAGAA TCGATTTACA 300  
 ACATTTCTTt GTCGATCGAT TATAAATTTT GTGAAATTC ATTTTATTGG GGaCCCAAG 360  
 ATTCCTGGTT GTTCaTTCTT TAAATGCGtA AATAACGGAT GTTCATCGTT CCCGTTcACA 420  
 40 GATATTTTAG CTAACACTGG AAATGTTACA CCAAAATTCT CACGACTAAT TTTCAAGATT 480  
 TCTTCATTAG AACCTGGTTG TCGATTGTCA AAATTATTAT TCGGAAAACt CAACACTACA 540  
 AACCACGAT CCTTATATTT TTGAAATAAA GTCTCTAGTT TTTTCAATTG TTCGCTATAT 600  
 45 ATACATTCTG TTGCAGTATT AACAACTAAA ATCACTTTAC CTCTAAATGC TTCTAATTTA 660  
 TAAGTTAAGC CTTTATAATC ACTTACTTCG ATATCATACA CATTTCATT ATTcATAAGA 720  
 CACCCCTACA CAGCCTTTTT TATATTGAAT ATGTTCTTTT TAGAATGTTc TGATAAAATA 780  
 50 AGTGCGCGTT TACACCGTGA ACACACATTA TATAGCGTGA TACATTTTTC GAGCACACGA 840  
 TAAATAATGT TCGAGTTTAT GTTGTTGCTC AACCTATCCG ATTTACCGTC TTTTTTCACC 900

55

CCGTTATAAC CCCTCATTAT AATCATCCTT ATTTTCTATT TTAAAAAGA CAATTAGACC 1020  
 GCTCTTTAAA CTATAGaTTA ATACTTAAGk TAAACTCATA CATACTGATA CCATACGTTA 1080  
 5 GATCTAACAA TTAAAAATTC GTTATAACTA TGGATTAAAG AGCTGCCCAA CTCATATAAT 1140  
 CCTTAAAAAC TTCACATGTG ATTGTkTATT AAGCCCTCCT TTATCnTATT AAATATCCTT 1200  
 10 ATAACCCTTT TAAAATTAAA CTGACACACT CATACTTGT TACAC 1245

(2) INFORMATION FOR SEQ ID NO: 335:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1399 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 335:

CTTGTAAGG TAGTTGTTCA TTAAATTAA AACAGTATGC TAAGTThGTG CTTATATTTT 60  
 GCAACTTCAA TCGACAGGC TCATCTTGAT ACATGAATGC CTCAATTTCA TCATGTGATA 120  
 25 ACTTTTTCOA AATATCAATT AAATGrATAC TAAAAACGAT AAATAACATA TCCCAATTTG 180  
 AATTCAGCCC TAGCGATTTT AATTGTGTTA AAATCTTATC TTTTGGAAA ATTTCGATTCC 240  
 TAATGTCTTG TATATCGTTA TCAGTCAAAG TTTCCCAATC TATATGTGAA TGAAGACCTA 300  
 30 AATAACACTT ATCCATTAAT AATTCATATA CCGTTAATGC AGAGACATCG AAACAACGTT 360  
 CTTCACTTAr AAAAACGCCA TCAACATCAA ATAAATTTT CTTCAACAATC CCCACTCCAT 420  
 35 TTCTGaAAAT TCAGaTATAA ATCATTCTAC TATTTGACTA AAAAAAGCGC AAACCCTATT 480  
 GAAGTaGATT TGCgCTTTAG CTGTThAAAT TTTATAAATG ThTTTCAATT TCATCAGCAA 540  
 CCTGCTGTAC GTGTGTACCG ACAATAACTT GAGTTGAATG TTTGCCATTA ACAGTAACAC 600  
 40 CAACTGCACC GGCGTTTTTA ATCTTCTGTT TATCAATAAT AGATGTGTCT TTAACTCTA 660  
 GACGCAACCT TGTTGCACAA TTGGTTAAAT TAACAATATT CTCTTGACCG CCTAAACCTT 720  
 CTAATATTTG TATAGCATGT TGATGATATT TACTTTGTTT AATATCATT TCACCAGGAG 780  
 45 CAATATTATC TTTTACAACT GTTGGGTCAA CTAATTCATT TTCACCTCTA CCAATCGTAT 840  
 TCAAGTTAAA TACTTGGATT ACTACACGGA AAATCACATA GTATAAGATG AAAAATACAA 900  
 50 CACCTTGAAC AAGCAACATC AATGGATGAT TTGATACTGG ATTAATTAGT GATAACACAT 960  
 AATCTATCAA ACCTGCACTA AATGAAAATC CAGCTGTCCA ATGGAATGTA GCTGCGATAA 1020  
 ATAAAGATAA TCCTGTTAAT AACGCATGAA CAACATATAA GATTGGCACA ACAAACATAA 1080

ACCAACCGTA AACTTGT TTTT TTCTGAGTAG TTTTAGCTGT ATGATACATT GcTAACGCAG 1200  
 CCGCTGGAAT ACCGAACATC ATGATTGGGA AGAATCCCGC TTGATAGCGT CCTGTAATAC 1260  
 5 CTTTTATAGC ATCTTTGCCA CTTTGAATT TACCAATATC ATTAATACCA ATCGTATCAA 1320  
 ACCAGAACAC ACTATTTCAGT GCATGATGTA ATCCTGTAGG AATTAATAAT CTATTGGCAA 1380  
 CACCATATAT GAAAGCTCC 1399

(2) INFORMATION FOR SEQ ID NO: 336:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1329 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 336:

TATAGTTnTA TTATTTAGCG AAGCATTAAAT ACTACCACCG GTTATAACAA ATGTATTTTG 60  
 CGTATTAAAT TGAATGGTAG GACCAATCAA AGTATATTCA ATCGCTGGAC CATCATTGTT 120  
 25 AATTAATGAC TGCGCAACCT TAAACTAAA TTGATCCATG GCACCTGCGC CTGAAAATCC 180  
 AATATGTTCA TAACCTATTC TTCCTAGATC TTGTACCGTT GAAAAGAGAC CTGGTTGTAA 240  
 AATCTTAATT GACATTTTCA ATCACCACCC AGTCATCAAC ATTAAAGTTG CCATCTGATA 300  
 30 TATCTCTTTC GATTTGTATA AATTTCTGTT CATCTATTGC ATAAAATTGT ATCCATTCTC 360  
 CTGCTTCGTA CATTGACATT GGTTACACGT CGCTGCTAAA TACTTTTAAc GGTGTGCGTC 420  
 CAATAATTTG CCATCCGCCA GGAGAATCTG ATGGATATAG TCCTGTTTGA TTATTCGCAA 480  
 35 TACCTACAGA ACCTGCATGA ATTTTAAACC TTGGCTGATT ACGTCTAGGT GTATGTAGTT 540  
 GTTCATCAAG TcCGCCTAAG TATGGAAATC CTGGCATAAA TCCTAGCATA TATATTAAAT 600  
 40 AAGGTTTACT TGTATGTTTT TCAATAACTT GCTCAACAGT TATTCGATTA TGCTTTGCTA 660  
 CTTCTTCAAT ATCTGGTCCA TATGTACCAC CATATTGAAC AGGTATTTTA ATAATACGAT 720  
 TGGTTTGATT CACAGCATGA ACATTTTTTTT CATTAAATTT GTTAAGTTCT AAATTTTCAA 780  
 45 TTAATTTAGA AGATGTTATA GCTTGTTTAT CAAAATATAT TAGAACTGCT CGATACGAAG 840  
 GGACAATATC TTGAATTTCT AATATTTCTT TTTCTCGTAT CCACCGTACC ATTGCTGTGA 900  
 CATTACGATA TGTCTCTTCG GATATTTTAT TTTCAAAATA AATCATAATT GTCTGCTCGT 960  
 50 TAATAAATCT TACATCCACT TTAAATCCCC CTTTGTATTG CAATAAACCA GTATTGAATA 1020  
 CCTTTTCATT GTATCATTGA GAAGCACAAG TTGTTTAATA AGTAATTCAA ATCGCATATA 1080

TTAATATTGT AACTCTTACA CTAATTTAGG TTCTGCTATC ATTCGGTCTG ATGGAAAATT 1200  
 TTTACTTTTC ATCTGTCCGA TTTTTTGATT TTGAATATAA AAAAGCACGA CCGAAGTATC 1260  
 5 ATTAACACAC TTCAATCGCG CAATTAAATA ATCTATTTGA TCATTTATTG GATATTAACA 1320  
 ATTTTACG 1329

## (2) INFORMATION FOR SEQ ID NO: 337:

10

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3421 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 337:

20 GAATAGTGCC ATTTGGAAAG ACATAAAGAA TCCAGAAGCA CCTTTAGCGA ATATGCCGTG 60  
 TTCATATAGA TTTGTGAAAC TGGCATGCCC GAATTGAGTT TTAAATGCAA ATAGAATCAT 120  
 GACGAAACCA ACTACTATTA AACCAATAAT TGTCGCTATT TTAATGATAG AGAACCAAAA 180  
 25 TTCTAATTCT CCGAAAAGTC TTGCGCTAAG TAGGTTGAAT GACATTAATA ACAGTACACA 240  
 AAATAGTGCA CTTATCCAGT TTGGAATTTT TGGGAACCAA AAGCTAACAT ATTTTGCCAC 300  
 AGCCGTTACT TCAGCCATAC CTGTAATAAT CCAACAGAAC CAGTATGTCC ATCCGGTAAC 360  
 30 AAATCCTGCA AAAGGCCCAA TATATGTATT GGTACATCT GCGAAAGATT TAAATTCAGT 420  
 ATTCTGTATA ATGATTTCTC CTAAACCTCG CATAAACATA AATAACATAA ATCCTATAAT 480  
 GATGTATGTT AATAGAATTG AAGGGCCGGT TAATGCAATC GTTTGACCAG CACCTAAGAA 540  
 35 TAAGCCTGTA CCAATTGCAC CGCCAATTGC AATTAAATTGT ATGTGGCGAT TGCTCAGTTC 600  
 CCTTTGTAAT TTTTCAGCCA TAATACATCT CCCTTAAATA TAGATATGTT TATTATGCAC 660  
 40 TTATATTGAG ATATATACAA TTATTTTCGG TAAAAATGTG TAAAATTCCA TGTTAATATA 720  
 CTTTGGTTTT TATAATCATA TATAATAACC AATTGAAAAT TTAATTCTAT TGTAAAATTC 780  
 ATGGATTATT CACATCTTGA AAAAGCTTTA ATGGTGCTAT TTGTGGCTAT TCTGTGACAT 840  
 45 TTACATAGAT TTACAAAAAA ATTGTTGCAC ATATAATGCC AGTtTTTATA TTTCACAAAC 900  
 GAAATGCGTT TAcTATAATA TTAGTTGAAA GCCATTTTCAT AAAGAAACAG TAAAGGGGAA 960  
 ATTTtATCaTA GCmGaATTAC AAAGAGGTTT AGAAGGGGTT ATCGCmGCGG AGACTAAAAT 1020  
 50 AAGTTCAATT ATTGAAAGTC AATTGACTTA TGCCGGCTAT GATATTGATG ATCTAGCTGA 1080  
 AAATGCGCAA TTTGAAGAAG TTATTTTCCT ATTATGGAAC TATAGATTGC CAAACGAAGA 1140

55



EP 0 786 519 A2

	TACACATTTT GAGGAGTATG TTACAGATCA CGTGCATCCA ATGACAGCAT TACGTACGTC	1260
	ATTATCATAT ATTGCACATT TCGATCCTGA TGCTGAAAAT GAATCAGATG AAAATCGTTA	1320
5	TGAAAGAGCA ATGCGTATAC AGGCTAAAGT AGCATCATTG GTTACAGCGT TTGCTCGAGT	1380
	AAGACAAGAT AAAGAACCAC TTAAGCCTAA TCCTGACTTA AGTTATGCGG CAAACTTCCT	1440
	ATATATGTTA CGTGGGGAAT TACCAACAGA TATAGAAGTA GAAGCCTTCA ATAAAGCACT	1500
10	TATTTTACAC GCTGATCATG AGTTGAACGC ATCTGCATTT ACGGCACGTT GTGCGGTATC	1560
	ATCATTGTCA GATATGTACT CAGGTATTGT AGCAGCCGTA GtTCTCTGAA AGGGCCATTA	1620
15	CATGGTGGTG CAAACGAACA AGTTATGACG ATGTTATCTG aGATTGGGTC AaTTGAAAAT	1680
	GTTGATGCTT ACTTAGATGA AAAATTGCT AATAAAGrTA AAGTAATGGG cTTCGGTCAT	1740
	CGTGTATATA AAGATGGTGm tCCTAGaGCG AAaTATTTaA GaGAAaTGAG CCGTCAaATT	1800
20	mCGAAAGACG CTGGTCGTGA AGAATTATTT GAAaTGTCAG TGAAaATGGA AAaTCGTATG	1860
	GCAGAAGAAA AAGGATTAAT TCCTAATGTT GATTTTTATA GTGCGAGTGT TTATCACTGT	1920
	ATGGAAATAC CTCATGACTT ATTCACGCCA ATCTTTGCTG TAAGTCGTTC TGCAGGATGG	1980
25	ATTGCTCATA TTTTAGAACA ATATAAAGAT AATAGAATTA TCGTCCTAG AGCGAAATAT	2040
	ATTGGCGAAA CGAATCGTAA GTATATCCCG CTTGrAGaAA GAAaMTAATC AATACAAATT	2100
	AAAAATGAAG ATGTAAAATT TGGAGGTAAA ATAACATGA CTGCAGAAAA AATTACTCAA	2160
30	GGAAGTGAAG GATTAAACGT ACCTAATGAA CCAATTATCC CATTTATTAT CGGTGATGGA	2220
	ATTGGACCGG ATATTTGGAA GGCAGCAAGC CGAGTTATAG ATGCTGCTGT TGAGAAaSCC	2280
35	TATAATGGCG AAAAACGCaT TGAATGGAAA GAAGTGCTAG CTGGCCAAAA AGCATTTGAT	2340
	ACAAGTGGTG AATGGTTACC TCAAGAAACA CTTGATACAA TTAAAGAATA TTTAATTGCT	2400
	GTTAAAGGAC CTTTAAACAC ACCAATTGGT GGTGGTATTA GATCATTAAA TGTGGCTTTA	2460
40	CGCCAAGAAT TAGATTTATT TACTTGCTTA AGACCGGTAC GTTGGTTTAA AGGAGTACCA	2520
	TCACCTGTTA AACGTCCACA AGATGTTGAT ATGTTATTT TCCGTGAAAA TACTGAAGAC	2580
	ATTTATGCTG GTATTGAATT TAAAGAAGGT ACAACAGAAG TTAAAAAGGT AATTGACTTC	2640
45	TTACAAAACG AAATGGGTGC GACAAACATT CGATTCCCAG AAACCTCAGG TATTGGTATT	2700
	AAACCAGTTT CTAAAGAAGG AACTGAGCGA TTAGTTAGAG CAGCTATACA ATATGCTATC	2760
50	GATAATAACC GTAAATCAGT TACTTTAGTT CATAAAGGTA ATATTATGAA ATTTACAGAA	2820
	GGCTCATTTA AGCAGTGGGG TTACGATTTA GCATTATCTG AATTTGGTGA TCAAGTATTC	2880
	ACTTGGCAAC AATATGACGA AATTGTTGAA AATGAAGGCA GAGATGCTGC TAATGCTGCT	2940
55		

TTACAACAAA TTTTAACTCG TCCAGCTGAG CATGATGTTG TAGCAACTAT GAACTTGAAT 3060  
 GGTGACTATA TTTCAGATGC TTTAGCTGCA CAAGTTGGTG GTATTGGTAT TGC GCCAGGT 3120  
 5 GCAAACATTA ATTATGAAAC AGGTCATGCT ATTTTTGAAG CAACACATGG TACAGCTCCA 3180  
 AAATATGCAG GTTTAAATAA AGTGAATCCA TCTTCAGTAA TTTTAAGTTC TGTATTAATG 3240  
 TTAGAACATT TAGGATGGCA AGAAGCGGCA GATAAGATTA CAGATTCAAT TGAAGATACA 3300  
 10 ATTGCTTCAA AAGTTGTTAC TTATGACTTT GCCCGTTTAA TGGaTGGtGC TGAAGAAGTT 3360  
 TCTACATCAG CATTTCGAGA TGAATTGATT GnAAATTTAA AATAAGCAGA ATAGAATTAG 3420  
 15 G 3421

## (2) INFORMATION FOR SEQ ID NO: 338:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3173 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 338:

CCCTnGATAC CCAAACGCC CTAATACCAC TGGCTAAACC TAATGGATAG TACCATTTAT 60  
 TTTCCAATAA ATAAhCCAAC TGCAATnGCT ATaACTCTAA ATATAATAGA GATAATCmCA 120  
 30 TTAATTGGAT TAAtACGCCA AATACTAGTA ATAATAGGCT AGATAATAAT CCACCTAAAA 180  
 AGTACTTTTT AATTCCAAAG AAAGCTAATA TCAATAATGC TGCCGGTGCA GATAATTGAA 240  
 AATCTAATCC TGGTATAATG GACGGTATTT TCAAACTGC CAAAATGGTT AAAATCGCAG 300  
 35 CAATGACACT AATTTGAGTA ATATCTTTTG ATGTCATACT AAAACCCCTA TACCGTTTCA 360  
 TAAACAACCTT GCTTCGGTGT GCTTTCTAAA AATGATATGT AATGATTTAA ATCAATACAA 420  
 40 TCGTCCACAA ATATTATTCT GCCTCCATAT CTCGTATTAA CTGGTTTAAT ATCAAATAAT 480  
 CGATGGTAAC CAATTTTAGC AGCGGCAAAA TAACCTGTCTG TATACGTTAA GTCATCGGAC 540  
 ACGCAAAGTT CTCCTTTGAC ATACGGATGC GCATTGATAC AACTAGCAAT TGCTAAGGCA 600  
 45 TCAGTCACTC TTTCATTAAG ATCACCTTTA TTATTTATAT CTTCAAACGA AAAATGTGTT 660  
 GCCCTAATCC CCCTTTGTCC AAATGAATCT AAACGTTTAC CAGATATAGC AGATAGAATA 720  
 ATAGCTCCTG TATAAACCGT TTCATTTTAA ATATATGTCA TCCCTTGATT TAGCGCTTGT 780  
 50 TCAGTGACAC CACATTCTTG TGTTAAATGT TGGAGATTG CTTTATCATC CTCAATAATT 840  
 TGTAATGCTT TTATTTGTTG AATCGGTTCC ATGATTTTTT GTATTTTTAT ATTGAGAAAA 900

55

	ATTTTGTCAA	ACTCACAAAT	CGTTTCAGCA	CCACTAATAT	GAACATCTTG	ATTGCTAGAA	1020
	CGCATTTTTA	TACTATACAT	GACGATCACC	TcAATCTTCT	TGaTGCAAAA	TTTCAAACAA	1080
5	CCTATCTATA	TCTTGTTTCA	TATGAAAATA	CGACAATGAT	ATTCTTAACA	TTGGCTTAGT	1140
	CACAGTtGGA	TACCTTAAAT	AACCTGTAAA	CACATGATGC	TTTAATAATG	TTTGATGAAT	1200
	GTTCTCAGCC	GCTTCTATGT	CATCAAACCTC	AATAAACTTA	ATCGGCGAGT	TTGaACTATT	1260
10	ATAATkAACA	TTGAGTGCTT	TTAACTTTTG	GTAAAAATAT	TTACTCAAAC	TATTTAATTT	1320
	AGTGCGTCTA	TCATCAGCAT	TTATTAACCTT	TTCAATGTTT	CTTTTTATAA	AATACAAATT	1380
15	ATAAATTGGC	AAACTACTTG	AGTAGATGAG	TGGTCTACCG	TGATTAATTA	aCATATCCtT	1440
	CaCATCaTTT	GaACTkaAAA	TcACACCCCC	GTATGCACCA	CATGCTTTAG	ATAAACTAGA	1500
	AGTGAGTATA	TCTACACCTT	GATAATTGCGA	GTAAtTCTCT	AttCCAAAAC	TATGTGAAAC	1560
20	ATCGAGTATC	AGTGTTGCGT	TAnATTTATG	CTTTAATGAG	ACTAATTGAC	CAATATCCAC	1620
	AACGTCGCCA	TTCGTTGAAA	ATACACTATC	AGATATGATT	ATTTTTGGTA	TATTTTGATT	1680
	AGGGTATTTT	TCTAACCTTT	TTTCTAAATC	AGCAATATCT	AAATGCTTAT	ATATCACTTT	1740
25	TTCTAAACCA	CTTAACTTAA	TACCGTCAAT	AATACTCGCA	TGATTTTCTT	GATCTGAAAA	1800
	CACGACACAA	TTTGATTTTT	TGAAAATATT	AAATAACGCC	AAATTAGCAT	CATAACCACT	1860
	ATTTAAGATA	GTACATGcAC	TATATCCGAG	CCAACCTGCT	AACATTGTTT	CAATTTCTTC	1920
30	ATAAGCTGTC	GAACCTCCAC	TAATTAATCT	TGAACCTGAT	AAGTGATAAC	TATACTTCCG	1980
	CATAAATCTT	TCGAAATCAT	CCTTATCAAA	CGCTATTTGA	CCTAATCCTA	AATAATCATT	2040
	AGATGTATAG	TTCGTACATC	TCTTATTTTC	TACTTCAATA	TACTGTCTAT	CTATATACCC	2100
35	TACCGATTTA	AGCGACCGAT	ATAACCCTTT	CTGTGTAAAT	AAATCAATTT	GCTCTTGAAA	2160
	CTTCATTCTT	GTTTTCCTTA	TTTTCACAAAG	TGTCATAATC	AATTTCAAAG	CCTAAATCAT	2220
40	TAATCATATC	GTAGTCTAAT	TGGTTCGGTT	GCCCACCAGT	AATTAGATAA	TnCACCGACA	2280
	AATATTGAAT	TCGCCGCTTT	TAATGCTAAT	GGCTGTAACG	AACGTAAGTT	GACCTCTCTT	2340
	CCTCCAGCAA	TACGAATTTT	TTTCGTAGGA	TTGATTAATC	GGAATAATGC	TACGATTCTT	2400
45	AAACATTTCA	TTGGTGTTAA	ATCATCCATG	CTTCCAAACT	TTGTGCCTTT	GATTGGATGC	2460
	AAAAAATTAA	TCGGAATACT	GTCGGCATCC	ATTTCTTTTA	AAGCAAATGC	CATATCAACA	2520
	ATATCTTGAT	TAGATTCTCC	CATACCACAA	ATCACGCCAG	AACATGGTGA	TATATTATTC	2580
50	GCTTTCATTA	GTTCTATCGT	ATCTGTTCTA	TCTTTATAAC	TATGCGTTGT	CACGACGTTA	2640
	TCATGGTAAT	TTTCACTTGT	ATTAATATTG	TGGTTATATC	TGTCTACACC	AGCTGACTTA	2700
55							

TGTGAGATT TAATCGTTCT TACAGTATTA CTAATATGAT CAACTTCTTT ATCGCTCGGT 2820  
 CCTCTACCAC TCATAACAAT ACAATATGTT CCAATATGAT TATCATGTGC CACCTTTGCT 2880  
 5 CCATCGATAA TTTGTTCTC TGGAATTAAA GCATATCGCT GTTTTGTTT AATATCTCGT 2940  
 GATTGTCCAC AGTACCCACA ATTTTCAGGA CATATACCAC TTTTAGCATT TAAAATCATG 3000  
 TTTAATTTTA CTTTTTTACC AAAATAATGT TTTCTTAAAA TGTACGCCTC ATTTAATAAA 3060  
 10 TCTAAGGTAT CAATATTAGT ATCCTCATAA ATTTTCAATA CAGTCTCTTT TGTtAATTGT 3120  
 tCCCCTTGTA ATATGCGTTT AGCCAAATTC ATATTAAACAC TTCCTATCTA AAA 3173

15 (2) INFORMATION FOR SEQ ID NO: 339:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1694 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

25 CGATTATCCA TTAATACAAC CCTAAGTAAA TGTATAAAAA TTATCTTCCA CAAACTTCAA 60  
 CAAAAGCCTA AATAAATTAC AGCAATTTAT CAAATATTGC TTACTTTGAT TTTATGAAAT 120  
 nACTTAATTC TAACACATAC TAAATCATCA TATACTAATT CGAAATCAAA TGCATTTAGA 180  
 30 GATAATCaAA ATGCGGAAAC ATCTCCaATA ATCAATAATC TATTCCCAAT AAATATGAAT 240  
 GTTCTCAACA ATACATTATT TATATCTCTT TACACTGTCA TCGACAAAAA CTAAATCTTT 300  
 CACTTTCAAT TTCGAACGTG GTTCTACGAC ATTTGCTGCT ATATCATTTA ATGGGATTAA 360  
 35 AACAAATGCA CGTTCATTCA TTCTCGGATG TGGCACCGAC AGTTTTGGTA AATCTATCAT 420  
 TTCTTCTCCA TACAACAAAA TATCCACATC TAAAGTTCTA GGACCCCATC GTTCCTTTCT 480  
 40 AATACGGTGT AAACATTCTT CTGTckTCAA ACAACATTCC AACAGTTGTA ATACTGTGAG 540  
 TGTTGTTTsA ATTTCAACAC ACAAATTTAA AAAGTTAGGT TGCTCAGTAT ACCCAACTGG 600  
 TGCTGTTTCA TAAATCGGAG AAATAtTAGA TACGTTAATA CCATCATATT CATTCaAAAT 660  
 45 CTTyATAGCA TCGTTTAAct GGCTTTCTCT ATCACCaaTa TTACTACCTA AyCCTAAGTA 720  
 TGCTTGAATC ATyTATTCTC CCTCACTATT TCGATACCTA CTCCATCATA ATGACCCGGA 780  
 ATCGGTGGGT TTTCTTTAGT GATTCTCACT TTCGTTTCCA TTACACGATT ATATTGTGAA 840  
 50 TTTATACGAT TTGCAATACG TTCAGCTAGA TGCTCAAGTA AATTAACGGC CTTACCTTCC 900  
 ATAATTGATT TAACCTCTTc GAACACTTCA CCATAATGAA CTGTATCAAT AACATTATCA 960

55

5  
 10  
 15  
 20

ATTTCATTTT	CAGCTGATAA	AGCACCATGA	TATCCATAAA	AGCGCATACC	TTTAAGAAAG	1080
ATTGTGTCTT	GCATTTTCAT	TCTCCTTTAA	AAAATCTATA	CCTTTAGCTA	ATTTAGCATT	1140
CAACTCGACA	TTATGAACGC	GTACTGCTCT	AACGCCTTTC	ATAATACCAT	ATGCAGTCGT	1200
AGCTGCAGTT	ACTTCATCTC	TTTCAACCGG	TGTTGTATCA	TAACCCATCA	TCTCTTTAGT	1260
GAAACGTTTC	CGGCTTGTCG	CTAATAAAAC	TGGATATTCT	GTTGCAACAA	GTTTCATCCAG	1320
TCTTGCCATA	ACTTCGGCTT	CTTCATTTCT	AGTTTATAGCG	AAACCTATAC	CTGGATCTAG	1380
CCAAATTTTA	TTTGAAGGTA	TACCAGCTAT	TTTAGCTTGA	TGTGCTTG TG	CTAACAAAGA	1440
TGTTAACATT	TCTTCGACAA	CCGGTTCATC	ACGATTACCA	TTTCCATTAT	GCATTAAAAT	1500
AATTTCCGCG	TCATAATTAG	CTACAATTTG	GAACATACGA	TGATCATACA	GACCGCCcAT	1560
tGATCATTAA	TCATATCAAC	GGCTAATTTT	AAACATGCTT	CAGCAACCTC	ACTTCGAAAT	1620
GTATCGACTG	AAATTTTTTA	CATCAnAACC	GACAATAGCT	TCAnCAACAG	TAATACTCTG	1680
TTCCATCTCT	TCTG					1694

## (2) INFORMATION FOR SEQ ID NO: 340:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1358 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

35  
 40  
 45  
 50  
 55

AGCATTTCTT	TTCTATAAAC	ATTTAATTGA	ACATTATTAA	GTACACTATT	ACTATAGTCA	60
CTATATTGAA	CaCATACCTC	ATTTAATTCT	AATAGCGGTT	CAGATTTGTA	CTTATTATCA	120
TTATTTGCAG	ATGTTTCATC	TATCCATTTT	TTCACTTTAA	ATTTAACATG	TTCAC TCATA	180
CAAACGTCAC	GTAAATTTCG	TAAGTTATCA	ATGGATTGGA	CATCTACTTC	TGCATATTTA	240
AGCGCTGTAC	AGTATAATGG	TTCACGTATG	CCTGCTTCTT	TAAGCTTAGA	TGATTTTAGC	300
AAATCACTAG	GCGTTGTATT	AGCGATGATT	TTTCCATCTT	TAAAAAGAAG	AACTCTATCA	360
AACGTATCAT	CTAATGATTG	TTCTAATCGA	TGTTTCGACAA	TAATCATCGT	TGACTTTGTT	420
TCTTCATGAA	TATTGTTTAA	CAATCTCAGC	GTTTCATGTC	CTGTCGCAGG	ATCTAAATTG	480
GCCAGTGGCT	CATCCAATAT	TAAAATAGGC	GTACGATGGA	TTAATATACC	ACCTAATGAA	540
ACGCGTTGTT	TTTGACCTCC	AGATAAATCT	TGCGGTCGGT	GATTTAAATG	TTCTATCATG	600
CCAAC TTTT	CAGCCCAATA	ACTTACATTT	TTCTTCATAT	CATCTTGTTT	AACACAATTA	660

TCTTGTAAAA CTGTACCAAC AACATTAGAT CTATCATGTA AACAACTAAC GGTTGCATCT 780  
 TGATTATTTA TATATAGTTC CCCAGTTATG TTACCTTTAG TTTTAAATGG AATTAATCCG 840  
 5 TTTATGCAAT TTGCAAAAGT CGATTACCA CTACCCGAAG CACCAACTAC TAATACTTTT 900  
 TCTCCTGGAT AAATATCAAC ATTTATATTC TGTAATGTAG GTGTTGCTTG ACTATGATAT 960  
 TGAAAACTAA AGTCTTTGAA CGAGATAATT GGTTCACTCA TGATATATCA TTACCTTTCT 1020  
 10 ATATTCATTT ACATATCTGA TTCAACAAAA TAACTATTCC TTACGTAAAC TACCTTTTTT 1080  
 AATTTGAGAT GAaGCATATG CTTTTAATAA TATTGTCCCA ATAATGCCAA CTGAAATAAT 1140  
 15 ATTTAATACT GCAGAGATAA CACCTTGTGT ATAAACCTTG TTAGCTGGTT CGTTATAAAT 1200  
 CAAAATATCT AATGTTGGTG CAATAAGTGC CCAGCAAATA ATATTGCAA TAATTTGACC 1260  
 GATATTAAAA TAAACCATCG ATTTCTAGA AAATCGGCCT GAAGAAAGAT TTAATTTTAG 1320  
 20 TCCAATCCAG CCATATAAAC AGCCTATAAT TCCCGAGC 1358

(2) INFORMATION FOR SEQ ID NO: 341:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4557 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

TAGAAGAATT GGAGAAAATG CTAATTCAAT TGTCAACATT CCATAGTTAT CATGATTTAG 60  
 AGTTTCTATT TGTGACACGT GAAGATGAAG TTGAAACATT GAAATGGGCA CGTTGGTTGC 120  
 35 CACATATGAC ATTGAGAnGG CAAAACATTA GAGGATTGTG TTACAATCAA CGAACGCGTG 180  
 ACCAAATTTT AACGTCAATT TATAGCATGA TTAAAGAACG TATCCAAGCT GTGctGaACG 240  
 40 CAGCAGAAGT AATGAGCAAA TTATTTTCAC ACCGCAATTA GTGTTTGTCA TTACAGATAT 300  
 GTCATTAATT ATTGATCATG TCATTTTAGA ATATGTAAAC CAAGATTAT CAGAATATGG 360  
 TATTCATTA ATCTTTGTTG AAGATGtGAT TGaAAGTTTG CCAGAGCATG TAGATACCAT 420  
 45 TATTGATATC AAGTCTCGTA CTGAAGGCGA ACTGATTACG AAAGAAAAAG AATTAGTTCA 480  
 ATTGAAATTT ACACCTGAAA ATATTGrTAA CGTCGATAAA GAATATATCG CGCGACGTTT 540  
 GGCGAATTTG ATACACGTCG AACATTTGAA AAATGCAATT CCTGATAGTA TTACATTTTT 600  
 50 AGAGATGTAT AACGTGAAAG AAGTAGATCA GCTTGATGTG GTTAATCGAT GGAGACAAAA 660  
 CGAAACATAC AAAACGATGG CAGTACCTTT AGGTGTAAGA GGTAAGATG ATATTTTATC 720

55

	AGGGAAATCT GAGATTATCC AATCATACAT TTTATCTTTA GCTATTAATT TTCACCCTCA	840
	TGAAGTTGCA TTCCTATTGA TTGACTATAA AGGTGGGGGT ATGGCGAACT TATTTAAAGA	900
5	TTTAGTCCAT TTAGTTGGTA CGATTACAAA CTTAGATGGC GATGAAGCGA TGCGTGCCTT	960
	AACATCAATC AAAGCCGAAT TGAGAAAACG TCAACGTTTA TTCGGAGAGC ATGATGTTAA	1020
	CCATATTAAT CAATACCATA AGTTATTTAA AGAAGGTATT GCGACAGAAC CAATGCCACA	1080
10	TTTATTCATT ATTTCCGATG AGTTTGCCGA ATTAAAATCA GAACAACCTG ATTTTATGAA	1140
	AGAACTTGTA TCAACGGCAC GTATTGGACG TTCGTTAGGT ATTCATTTAA TACTTGCGAC	1200
15	ACAAAAACCA TCGGGTGTG TTGaTGACCA AATTTGGTCT AACTCTAAAT TTAAGTTGGC	1260
	ATTTAAAGTA CAAGATAGAC AAGACAGTAA TGAAATTTTA AAAACACCAG ATGCAGCAGA	1320
	CATTACmTTA CCaGgTCGTG CGTATTTACA AGTTGGTAaT AATGAmATTT ATGAATTATt	1380
20	CCAATCTGCA TGGAGTGGTG CAACATATGA CATCGAAGGC GATAAATTAG AAGTTGAAGA	1440
	TAAGACGATT TACATGATTA ATGACTATGG TCAACTTCAA GCAATCAACA AAGACTTGAG	1500
	TGGACTTGAA GATGAAGAAA CGAAAGAAAA TCAAAGTGGG TTAGAAGCGG TCATAGATCA	1560
25	TATCGAATCT ATTACAACAC GATTAGAAAT CGAAGAAGTT AAGCGTCCAT GGCTACCACC	1620
	ATTGCCAGAA AATGTATATC ArGAAGATTT AGTAGAAACa GATTTcAGAA AATTATGGTC	1680
	AGATGATGCA AAAGAAGTGG AATTAACATT AGGACTTAAA GACGTACCAG AAGAACAATA	1740
30	TCAAGGACCG ATGGTATTGC AATTGAAAAA AGCTGGGCAC ATCGCGTTAA TCGGAAGTCC	1800
	AGGATATGGT AGAACAACGT TCTTACACAA CATTATTTTC GATGTTGCAA GACACCATCG	1860
35	TCCTGATCAA GCACACATGT ACTTGTTTGA TTTCCGTACC AATGGTTTGA TGCCAGTTAC	1920
	AGACATACCA CATGTCGCTG ATTACTTTAC AGTAGATCAA GAAGACAAGA TTGCTAAGGC	1980
	GATACGTATA TTTAATGATG AAATTGATCG TCGTAAGAAG ATTTTAAGTC AGTATCGTGT	2040
40	CACTAGTATT TCTGAATATC GAAAATTAAC TGGTGAAACA ATTCCGCATG TCTTTATTCT	2100
	TATTGATAAC TTTGACGCAG TAAAAGATTC ACCTTTCCAA GAAGTTTTTG AAAATATGAT	2160
	GATTTAAATG ACGCGTGAAG GGCTAGCATT AGACATGCAA GTAACCTTAA CTGCTTCAAG	2220
45	AGCTAACGCT ATGAAAACAC CAATGTACAT TAATATGAAA ACGCGTATCG CAATGTTTTT	2280
	ATATGATAAA TCAGAGGTGT CGAACGTAGT AGGACAGCAA AAATTTGCGG TTAAAGATGT	2340
	TGTGGGTGCG GCATTGTTAA GTAGTGATGA CAACGTATCA TTCCATATTG GCCAACCATT	2400
50	TAAACATGAT GAGACCAAAT CATATAATGA TCAAATTAAT GATGAAGTAT CGGCGATGAC	2460
	AGAATTTTAT AAAGGTGAAC ACCAAATGAT ATtCCTATGA TGCCAGATGA AATTAAATAT	2520

55

## EP 0 786 519 A2

	GGATTAGATT ATGAAGGTGT TACACTACAA AAAATTAAAT TAACTGAACC AGCAATGATT	2640
	TCATCAGAAA ATCCGAGAGA AATTGCGCAT ATTGCTGAAA TTATGATGAA AGAAATTGAC	2700
5	ATATTAAATG AAAAATATGC GATTTGTATC GCAGACTCAA GTGGAGAGTT TAAAGCTTAT	2760
	AGGCATCAAG TGGCTAACTT TGCCGAAGAA AGAGAAGACA TTAAAGCGAT TCATCAACTA	2820
	ATGATTGAAG ACTTAAAGCA AAGAGAAATG GACGGCCCAT TTGAAAAAGA TTCACTTTAT	2880
10	ATTATCAATG ATTTTAAAC ATTTATTGAT TGCACGTATA TTCCGGAAGA TGATGTTAAA	2940
	AAGCTTATTA CAAAAGGACC AGAACTTGGC TTGAACATTT TATTTGTCGG CATTCATAAA	3000
	GAATTAATAG ATGCTTATGA TAAACAGATT GATGTTGCAC GTAAAATGAT TAACCAATTT	3060
15	AGTATAGGTA TTCGTATTTC AGACCAACAA TTCTTTAAAT TTAGATTTAT TCAACGAGAA	3120
	CCTGTTATTA AAGAAAATGA AGCATATATG GTCGCAAACC AAGCTTATCA AAAGATTAGA	3180
20	TGGTTTAAAT AGCAATGAAT TAAATAGGAG GGAGGTATGT TATGAATTTT AATGATATTG	3240
	AAACAATGGT TAAGTCGAAA TTAAAGATA TTAAAAAGCA TGCTGAAGAG ATTGCGCATG	3300
	AAATGAAGT TCGTTCTGGA TATTTAAGAA AAGCTGAACA ATATAAGCGA TTAGAATTTA	3360
25	ATTTGAGTTT TGCACTAGAT GATATTGAAA GCACAGCAAA GGACGTACAA ACTGCAAAAT	3420
	CTAGTGCTAA TAAGGACAGT GTAAGTGTTA AGGGAAGGC GCCCAATACG TTATATATTG	3480
	AAAAAGAAA TTTGATGAAA CAAAAGCTTG AAATGTTGGG TGAAGATATC GATAAAAATA	3540
30	AAGAATCCCT CAAAAAGCT AAGGAAATTG CTGGCGAAAA GGCAAGTGAA TATTTTAATA	3600
	AAGCAATGAA TTAATATTGA GGTGAAGATA TGGGTGGATA TAAAGGTATT AAAGCAGATG	3660
	GTGGCAAGGT TGATCAAGCG AAACAATTAG CGGCAAAAAC AGCTAAAGAT ATTGAAGCAT	3720
35	GTCAAAAGCA AACGCAACAG CTCGCTGAGT ATATCGAAGG TAGTGATTGG GAAGGACAGT	3780
	TCGCCAATAA GGTGAAAGAT GTGTTACTCA TTATGGCAAA GTTTCAAGAA GAATTAGTAC	3840
40	AACCGATGGC TGACCATCAA AAAGCAATTG ATAAGTTAAG TCAAAATCTA GCGAAATACG	3900
	ATACATTATC AATTAAGCAA GGGCTTGATA GGGTGAACCC ATGATGAAAG ATGTTAAGCG	3960
	AATAGATTAT TTTTCTTACG AAGAATTAAC AATTTTAGGT GGTAGTAAAT TGCCTCTCGT	4020
45	AAATTTTGAA TTGTTTGATC CATCAAATTT TGAAGAAGCT AAAGCTGCTT TAATTGAAAA	4080
	GGAATTAGTA ACAGAGAATG ACAAGTTAAC TGATGCAGGT TTAAAGTGG CTACATTAGT	4140
	CAGAGAGTAT ATTAGCGCCA TTGTAAATAT TCGAATTAAT GATATGTATT TTGCACCATT	4200
50	TAGCTATGAA AAAGATGAAT ATATTTTGTT AAGCCGTTT AAAAATAATG GGTTCAAAT	4260
	ACGAATTATC AATAAAGACA TTGCATGGTG GTCGATTGTA CAATCATATC CTTTATTGAT	4320
55		



CTTAAATAAT GAAAGTATCG ATACGATTGG GCGTGTTTTA GAAATTGAAA TATACAATCA 4440  
TCAAGGTGAC CCTCAACAAA GTTTATATAA CATTTATGAA CAAAATGATT TGTATTTCAT 4500  
5 TCGATACCCA TTAAAGATA AAGTGCTGAA TGTTTCATATT GGTGTCATTA ATACATT 4557

(2) INFORMATION FOR SEQ ID NO: 342:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3931 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

TTGAGTGACT TTATTGAAGC GCGTGTAAGG GAAATATTCT TCGAAGTATT TGATGTTTTA 60  
20 CAAGATTTAG GATTAACAAA AGTAAATGGT GGGTTTATTG TAACTGGTGG ATCTGCAAAAC 120  
TTACTTGCGC TAAAAGAATT ATTATCAGAT ATGGTAAGTG AAAAAGTTAG AATTCACACG 180  
CCATCACAAA TGGGAATTAG AAAACCTGAA TTTTCTTCAG CAATTTCTAC AATTTCTAGT 240  
25 AGTATCGCTT TTGATGAGTT ATTAGATTAT GTTACAATTA ATTATCATGa TAATGAAGAA 300  
ACTGAAGAAG ATGTTATTGA TGTGAAAGAC AAAGATAACG AATCTAAAT AGGCGGaTTT 360  
GaTTGGTTTA AACGTAAAC AAACAAAAAA GATACTCATG aAAATGAAGT AGAGTCAACA 420  
30 GATGAAGAAA TTTATCAATC AGAAGATAAT CATCAGGAAC ATAAACAGAA TCATGaACAT 480  
GTTCAAGACA AAGATAAAGA TAAAGAAGAA AGTAAATTCA AAAAATAAT GAAATCTCTA 540  
TTTGAATGAT TATTGGCCAA TAAACTAGG AGGAAATTTA AATGTTAGAA TTTGAACAAG 600  
35 GATTTAATCA TTTAGCGACT TTAAAGGTCA TTGGTGTAGG TGGTGGCGGT AACACGCCG 660  
TAAACCGAAT GATTGACCAC GGAATGAATA ATGTTGAATT TATCGCTATC AACACAGACG 720  
40 GTCAAGCTTT AAACCTATCT AAAGCTGAAT CTAAATCCA AATCGGTGAA AAATTAACAC 780  
GTGTTTATAG AGCAGGAGCT AATCCTGAAA TCGGTAAAAA AGCTGCAGAG GAATCTCGTG 840  
AACAAATTGA AGATGCAATC CAAGGTGCAG ACATGGTATT TGTTACTTCT GGTATGGGTG 900  
45 GCGGAAGTGG TACTGGTGCA GCACAGTCG TTGCTAAAT TGCAAAAGAA ATGGGCGCAT 960  
TAACTGTTGG TGTGTAACT CGTCCATTTA GTTTTGAAGG ACGTAAACGT CAAACTCAAG 1020  
CTGCTGCTGG AGTAGAAGCT ATGAAAGCTG CAGTAGATAC ATTAATCGTT ATACCAAATG 1080  
50 ACCGTTTATT AGATATCGTT GACAAATCTA CGCCAATGAT GGAAGCATTT AAAGAAGCTG 1140  
ACAACGTGTT ACGCCAAGGT GTACAAGGTA TCTCAGACTT AATCGCTGTT TCTGGTGAAG 1200

## EP 0 786 519 A2

	GTATTGGTGT TTCTTCTGGT GAAAATAGAG CGGTAGAAGC TGCTAAAAAA GCAATCTCTT	1320
	CTCCATTACT TGAAACATCT ATCGTTGGTG CACAAGGTGT GCTTATGAAT ATTACTGGTG	1380
5	GCGAGTCATT GTCATTATTT GAAGCACAAG AGGCTGCTGA TATTGTCCAA GATGCTGCAG	1440
	ATGAAGACGT TAATATGATT TTCGGTACAG TTATTAATCC TGAATTACAA GATGAGATTG	1500
	TTGTAACAGT TATTGCAACT GGTTTTGATG ACAAACCAAC ATCACATGGT CGTAAATCTG	1560
10	GTAGCACTGG ATTCGGAACA AGCGTAAATA CTTCTAGCAA TGCAACTTCT AAAGATGAAT	1620
	CATTCAC TTC AAATTCATCA AATGCACAAG CAACTGATAG TGTAAGTGAA AGAACACATA	1680
15	CAACTAAAGA AGATGATATT CCTAGCTTCA TTAGAAATAG AGAAGAAAGA CGTTCAAGAA	1740
	GAACAAGACG TTAATCGGTT AATATATATA CACAAATAAT TCAACACAAA TCATCAGATA	1800
	ACATATCTGA TGATTTTTTT ACTAATTTTT AGaACATGTA GAAGGACATT TAAGTTTTTC	1860
20	aAAGTTATTA AAAGTGTTTA AGTATCGTGT GAAAATTAAG TCaAAAATTA TTTGCGCAAC	1920
	ATTTTAACTT TAAACATAAA TGTTATATTA TATAATTATT AACTTTGTAC AGTTAGACGA	1980
	AGATAATTTA AATGAAATGA TGGTGACGAT CGAGTGAATG ATAATTTTAA AAAGCAACCG	2040
25	CATCATTTAA TATATGAAGA GTTATTACAA CAAGGTATTA CTCTAGGTAT TACAACTAGA	2100
	GGAGATGGTT TAAGTGACTA TCCTAAAAAT GCTTTTAATA TGGCGAGATA TATTGATGAT	2160
	CGCCCATATA ATATTACTCA ACATCAATTG CAATTAGCTG AAGAAATTGC GTTTGATAGA	2220
30	AAAAATTGGG TGTTCCTCAT TCAAACACAT GAAAATAAAG TCGCTTGAT TACAAAGGAT	2280
	GATATAGGCA CAAATATAGA CACTTTAACT GATGCGCTTC ATGGTATTGA TGCGATGTAC	2340
	ACATATGATA GTAATGTCTT ATTAACGATG TGTATTGCAG ACTGTGTACC AGTATATTTT	2400
35	TATAGTACAA AACATCATTT TATTGCATTG GCGCATGCAG GTTGGCGTGG TACCTATACT	2460
	GAAATTGTAA AAGAAGTGCT AAAACATGTG AACTTTGATT TGAAAGACTT ACATGTCGTT	2520
40	ATTGGACCAT CTACATCATC AAGTTATGAA ATTAATGATG ATATTAAAA TAAATTTGAA	2580
	ACATTGCCAA TTGATAGTGC CAACTATATT GAAACTAGAG GACGAGATCG TCATGGTATT	2640
	GATTTGAAAA AAGCCAATGC TGCATTATTA ATTTATTATG GTGTTCTTAA AGAAAAATATT	2700
45	TATACGACAG CGTATGCTAC ATCTGAACAT TTAGAATTAT TTTCTCTTA TCGATTAGAA	2760
	AAAGGTCAAA CAGGACGCAT GTTAGCATTG ATTGGTCAAC AGTAAACAAG GAGGAGATAT	2820
	GTTTGCGTGT GAAAGATAAT TTACAACAAA TCTCAACACA AATTAATGAC AAAAGTGAAA	2880
50	AAAATAATTT TTCAACAAAA CCAAACGTGA TTGCAGTTAC AAAATATGTT ACAATAGAGC	2940
	GAGCTAAAGA AGCGTATGAG GCTGGAATAA GACATTTTGG TGAGAATAGA TTGGAAGGCT	3000
55		

AATCTCGAAA AGTTAAGGAC GTTATAAACG ACGTAGATTA TTTCCATGCT TTAGATCGAT 3120  
 TGAGCTTAGC CAAAGAAATT AACAAACGTG CAGAACATAA AATTAAATGT TTCTTGCAAG 3180  
 5 TGAACGTTTC GGGAGAAGCT TCTAAACATG GTATTGCTTT AGAAGATGTT GATCAGTTTA 3240  
 TAGATGATCT TAAAAAATAT GACAAAATCG AAATTGTAGG TTTAATGACG ATGGCACCAT 3300  
 TGACAGATGA TGAAGCATAT ATTAGATCGT TATTTAAACA GTTACGTTTG AAAAAAGAAG 3360  
 10 AAATACAACG ACTCAATTTA GAATATGCGC CTTGTGATGA ATTATCAATG GGAATGAGTA 3420  
 ATGACTATCT TATTGCAGTT GAAGAAGGTG CGACGTTTGT TAGAATTGGG ACTAAACTTG 3480  
 TAGGAGAAGA GGAGTGAGCC ACTTGGCTTT AAAAGATTTA TTTAGTGGAT TTTTGTAAAT 3540  
 15 AGATGATGAA GAGGAAGTAG AAGTACCTGA CAAACAACAA CAGGTAAATG AAGCGCCAGC 3600  
 AAAAGAGCAG TCACAACAAA CAACAAAACA AAACGCAATC AAATCAGTCC CTCAAAAATC 3660  
 20 TGCATCAAGA TATACAACAA CGTCAGAAGA AAGGAATAAC CGTATGTCTA ATTATTCAAA 3720  
 AAATAATTCA CGTAATGTTG TAACTATGAA CAATGCTACA CCAAACAATG CATCACAAGA 3780  
 AAGTTCAAAA ATGTGTTTAT TCGAACCACG TGTTTTTTCA GATACACAAG ATATTGCTGA 3840  
 25 TGAGCTTAAA AACC GCCGTG CGACACTTGT CAATTTACAA CGTATTGATA AAGTATCAGC 3900  
 GAAAAGAATT ATTGATTTTT TAAGCGGTAC T 3931

## (2) INFORMATION FOR SEQ ID NO: 343:

- 30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3150 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

40 AATTGTCGGG GGACTCTTAG GTTTTGT CAT GCAAAGAACA AGATTTTGTT TAACAGGTGG 60  
 CTTTCGAGAT ATGTATGTGC AAAAGAATAA TAAGATGTTT TATGCATTAT TAATCGCTAT 120  
 TACTATTCAA AGTATAGGAT TATTGATTTT GACGGCAACA GATATTTTAC AAATTCCTGC 180  
 45 ACATAGTTTT CCAATATTGG GAACAATTAT AGGTTCTTTT ATTTTGGAA TTGGAATAGT 240  
 ATTGGCTGGA GGATGTGCAA CAGGLACTTG GTATCGCGCT GGTGAAGGGC TAATTGGTAG 300  
 TTGGATTGCA TTAGTATTAT ATGCTGTTAC TGCAGCAATC ACTAAAACAG GGATTTTAAA 360  
 50 GCCAGTAATG GATAAAATTA ATCAACCAAC GAATGTAAAT AGTGATATGT CTCAAACAAC 420  
 TGGCATTCCG TTTTGGGGAT TAGTCGTTAT ATTA ACTATA ATCACCATT TTTAGTTGT 480

55

	AGGTATTAGA TATTACCTTT TCGAAAAACG ATACCATCCA TTTATTGCAG CAATTGTAAT	600
	TGGACTTATC GCACTCTTAG CTTGGCCAAT GAGTGCATCA ACTGGAAGAA ATGACGGTTT	660
5	AGGTATAACA ACGCCTTCAG CAAATTTAGT ACACTTTTTTG ATTACAGGTG AAACTAAATT	720
	TATTGATTGG GGTGTCTTTT TAGTTCTAGG AATTTTCATT GGTTCATATA TTGCAGCTAG	780
	AGGATCAAGA GAATTTAAAT GCGGATTGCC AGACAAGATT ACAATACGAA ACAGTGCCAT	840
10	TGGTGGCATA TGTATGGGAT TTGGTGCCTC AGTTGCTGGT GGTGTCTCTA TCGGTAACGG	900
	TTTGGTTGAA ACGGCAACGA TGA CTCTGGCA AGGATGGATT GCGCTAGCAT GCGATGATAG	960
15	TTGGTGTATG GACAATGAGT CATTTTATCT TTGTTCTGCC AATGAAAAAA GTACACCAAC	1020
	AATCTGCAAA GGTAAACAG CAAACGCAA TAGTATAGAA GATTATTATG CAAATGATGT	1080
	TGATCAAATA AAAGTGATTG GAAAAGGAGA AATAATTATG ATACACGAAT TAGGTACAGT	1140
20	AGGAATGGTA TGTCCATTTT CGTTAATTGA AGCGCAAAAG AAAATGGCAA CATTGCAATC	1200
	TGGAGATGAA TTA AAAAATTG ATTTTGATTG CACGCAAGCG ACGGAAGCCA TTCCAAATTG	1260
	GGCTGCAGAA AATGGTTATC CTGTAACAAA CTATGAACAA ATTGATAATG CTTCATGGAC	1320
25	AATTACAATT CAAAAAGTTT AACGTTATCA TTTTAACAAT AAAATAGATA TTAGATTCTA	1380
	TGGCTACTTC CGCTAATTTA AAAGTGAGTA AGTAGTCTTT TTTTTTTTAG TTCATGAAAT	1440
	CATTTTATA TAGTGTGGCA CATTTTATTC CAAAAGATGT AATAAACTT AACGCATTTT	1500
30	TGCTTTTTAT AAATTGTCAG ATTATTATGA AAAAAAGGGA GTGGTAAGTA TGAATCTTAA	1560
	CGATACGATA TTTATGTTTT TGTGTACATT ATTAGTTTGG TTAATGACAC CAGGATTAAG	1620
35	TTTATTTTAT GGTGGGTTAG TTCAATCTAA AAATGCGCTT AATACTGTCA TGCAAAGTAT	1680
	GGCAGCAATT GTGCTTGTTA CATTTGTATG GATAACAGTT GGTTTTACAA TTAGTTTTGG	1740
	GAATGGGAAT TTATGGTTCC GAAATTGGGA ATATACTTTT CTTAATCATG TAGGTTTTGC	1800
40	GACTCAAGAA GATATTAGCC CACATATTCC TTTCGCTTTG TTTATGTTAT TTCAAATGAT	1860
	GTTTTGTACG ATTGCAATTT CTATTTTATC TGGTTCAATC GCTGAGAAAA TGAAGTTTAT	1920
	TCCTTATTTA TTATTCGTAG TAATATGGAC TGCTCTTGTA TACAGTCCAG TAGCACATTG	1980
45	GGTTTGGGGC GCGGTTGGA TTAACAACT CGGTGTATTA GATTTGCTG GAGGTACGGT	2040
	TGTTCAATTT ACATCAGGTG TTTCTGGTTT AGTATTAGCT ATTATGATTG GAAAAGGAAA	2100
	CAAACATTCT GAATCAACAC CACATAATCT TATCATIACG TTGA1TGGCG GTATATTCGT	2160
50	GTGGATTGGT TGGTATGGAT TTAATGTAGG TAGTGCTTTT ACATTTGATA ATATTGCGAT	2220
	GCTTGCAATTT ACAAATACTG TCATTTTCAGC CAGTGCAGGT GCTATAGGTT GGTAAATTTT	2280
55		

5     ATTAGTTGTC ATTACTCCTG CAGCAGGATA TGTAAACATAT CTTAGTGCAA CAATAATGGC     2400  
       TTTAATAGGA GGTATCTGTT GTTATATTGT CATTAATTAC ATCAAGGTAA AACTAAAATA     2460  
       TCATGATGCA TTAGATGCAT TTGGTATTCA TGGTGTGGT GGTATTATTG GTGCTGTTTT     2520  
       AACAGCAGTT TTCCAAAGTA AAAAAGCCAA TCCTGACATT GAGAATGGCT TTATTTTATAC     2580  
 10     TGGTGACATA CATATTATAC TTGTACAAAT ATTATGTGTA ACAGCAGTTG TAATTTTATAG     2640  
       TATAGTCATG ACGTTTATTA TTGCGAAAGT AATTAAATTa ATTACACCAT TATCTGTTAC     2700  
       GGAACAAGAA ACGAATATAG GATTAGACAA GATTGTTCaC GGTGAACATG CTTACTTTGA     2760  
 15     AGGTGAGCTA AATAGATTCA ATAAACATAT TCGATATTAG AATATATTTA CATAGAATAT     2820  
       TCATTGTCCT GACATTTAAC TAAAGGTTGA TGTGGGACA TTTTGTTATA CAAAAGTTTT     2880  
       ATTTTGAAAT CTTTTTATGA AAGAAGCAGA AATATTATTT AAAGCGGTTA CACATATGCT     2940  
 20     AAAATAAGGC TAAGTGTCAC AAATAATGAT AGGTGAATAA GTATGAAAAA TATATCTGAT     3000  
       ATTGCCAAAT TGGCAGGCGT TTCAAAAAGT ACAGTATCTA GATTTTTTAA TAATGGATCT     3060  
       GTCAGTAAAA AAACAAGTGA AAAATTAACA AGAATTATAG CAGAACATGA CTATCAACCG     3120  
 25     AATCAATTTG CTCAAAGTTT AAGAGCGAGA     3150

(2) INFORMATION FOR SEQ ID NO: 344:

30     (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 3719 base pairs  
       (B) TYPE: nucleic acid  
       (C) STRANDEDNESS: double  
       (D) TOPOLOGY: linear

35     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

40     GTTATAGTGA AATTGACTCA TCACATTTCA CAGACCGTGA CAAACGCGTT ATTAGACGTG     60  
       ATCATGTTAA AGAAGCACAA AGCTTAGTAG AGAACTATAA AGATACACAA AGTGCTGATG     120  
       CTAGGATGAA AGCCAAACAA AAAGTTAACA CATTAAGCAA ACCGCATCAA AACTATTTCA     180  
       ATAAACAAAT TGATAAGGTT TATAATGGAT TACAACGCTA ATCCAAAGTA AATTATAAGT     240  
 45     TATACATCTC GTTTTTAAAT GACAATTTAT CCCCCTAAAT ATTATAAATA ATCTTTTCAA     300  
       ATTCCACATA GATATAGAGA CACTAATAAA CCTCTTTGTC TCGATATGAT AGTCTGCAAC     360  
       GATTGATGTT GTAGGCTTTT TAATTTTACA AATAAGGCTA AATATATAAG TTCTGGCACC     420  
 50     TAAATATAG AAAATACATA AAAGTAAGTA TAGTTATTTT ATTATAATTA TTAAATTTT     480  
       ATTAATTAAT TGTAAAAATG TATAATTATA ATTAATTAAC GTTTAATATT AAAATTAAC     540

55

	ATCGTTTCAA TATTACTTAT AGGGATGGCT ATCAGTAATG TTTCGAAAGG GCAATACGCA	660
	AAGAGGTTTT TCTATTTTCGC TACTAGTTGT TTAGTGTTAa CTTTAGTTGT AGTTTCAAGT	720
5	CTAAGTAGCT CAGCAAATGC ATCACAAACA GATAATGGCG TAAATAGAAG TGGTTCTGAA	780
	GATCCAACAG TATATAGTGC AcTTnCAACT AAAAAATTAC ATAAAGAACC TGCGACATTA	840
10	ATTAAAGCGA TTGATGGTGA TACGGTTAAA TTAATGTACA AAGGTCAACC AATGACATTC	900
	AGACTATTAT TGGTTGATAC ACCTGAAACA AAGCATCCTA AAAAAGGTGT AGAGAAATAT	960
	GGTCCTGAAG CAAGTGCATT TACGAAAAAA ATGGTAGAAA ATGCAAAGAA AATTGAAGTC	1020
15	GAGTTTGACA AAGGTCAAAG AACTGATAAA TATGGACGTG GCTTAGCGTA TATTTATGCT	1080
	GATGGAAAAA TGGTAAACGA AGCTTTAGTT CGTCAAGGCT TGGCTAAAGT TGCTTATGTT	1140
	TATAAACCTA ACAATACACA TGAACAACCT TTAAGAAAAA GTGAAGCACA AGCAAAAAAA	1200
20	GAGAAATTAA ATATTTGGAG CGAAGACAAC GCTGATTCAG GTCAATAATG CTCATTGTAA	1260
	AAGTGTCACT GCTGCTAGTG GCACTTTTAT AATTTTTAGA TCACGATATG ATTTATTATC	1320
	AATTCAGAAT TAAAAAAGTA AATAGTATCA AAAGTAAGTG TATTTAATAT TAGAAAAATA	1380
25	AAATTTTAAA TTTAGTATTA AAATGGAATG TTACTATATA GTTCAATGTG TATTATCACA	1440
	GAAAAATAAA TAATGCTTTA CTTCTATATT TAAAAGTGTA TAATGAAAGT TAAGTAATAA	1500
	AGAGCGTGAA GAAAAATGTG AGTTATTTAT ATAGAATATT CTCCTTTTCA TTTATGAATT	1560
30	TGTTACAAAA TATTTAGTGC AAAAGCACGA cGGAGGTATT CAATATGaAT AACGGTACAG	1620
	TTAAATGGTT TAATGCAGAA AAAGGTTTTG GTTTCATCGa AAGAGAAGAT GGTAGCGACG	1680
35	TATTCgTACA CTTCTcAGCA ATCGCTGAAG ATGGATACAA ATCATTAGAA GAAGGCCAAA	1740
	AAGTTGAATT CGACATCGTT GAAGGCGACC GTGGCGAGCA AGCTGCAAAC GTAGTTAAAA	1800
	TGTAATTTTA ACTTATTCAA ACAGTCCTTA CTATAGGGCT GTTTTTTTAT GCTTTAAATC	1860
40	GATAACAGTT GGTGTGGTAA AAGCACTAGC CGTTATTTTT TTGTCCAATA AATTTAGTTG	1920
	GAGATTTAAC AATATATAAT GGTTCTAAAA TAAATCGAAC TGATGGAAAA GTTTTTTACT	1980
	TTTCATCTGT CCGACTTTTG ATTTTGAATA TAAAAAGCG CCAATACAGA ACTTTAATAA	2040
45	TGACGAGAAT TAAAGTCTGT ATATGGCGAT AACAAGAAGT AATGTTAAAC ACTCAAATG	2100
	TTTAACAATA ATAGGATACC ACATCGCATA ATATCTTACT ACTTAATTAA TAATTTAACT	2160
50	AATCAACTTT TTGTTAATTT TTTATTAAGA CTGATTAATT ATTGAGAATA TTTATTGTTT	2220
	TTAAATCTC ATAATAATTC AGTAATCTTG TTTTCATTTA AAAGGCGAAA CATTAAAAATA	2280
55	ATTAAATAAA AATATTGCGT TTAATTTACA GCGTCAAATA TACTTATTC TAATGCTTTG	2340

ATTTCTGTGA GTATTTGGAA GCTACCATTA GGCAACGGTT TAACAATAGA CAATTGCTTT 2460  
 TCCGCTTGTT GTATTAAAAA AGGTTTTGTA GATTGATTAT TAATATGCCA TTCACTCATG 2520  
 5 TATGTTTTTC ACTCCTGCTT TAAAATAGGG TTAGAAAGTT TATAGTTGAG ACATTCATGT 2580  
 TCAACCAAAA TTTTGTTCGA ATTCAATAAA TGTCTTGTTT AAAATAGAAA TATTGTAAAT 2640  
 GTTATCGTCC AAAACTTCAC CAGTTAAGTA TTTGTTTTGA ATTAAAAATT GGCAGTTAGT 2700  
 10 TAAGAAGTCT TGATAATCAC GATCGCAAAA ATAGTTTTCA CGTGCATCTT TAGCATCGCC 2760  
 AAAAAAGTTA GCGACTGTTT CTGTTTCTCC TTTATTGCGA CGTTCAATAT ATAATTTGTA 2820  
 AAATTTAGCT ATTGTATACT TTTGTTCTTT AGTTAGTTCA TTCAAAATAT TGGGCCTCCT 2880  
 15 GAAATATCAT TTGTAATCTA TACCCAATTT ATTGCAAAAC AAAACTAAT TTAACTATTT 2940  
 GATGAAACTG TGTTAATAAg CTTTAACAAG CCTTAGTTTG TATGGATCTA TAAATTTATC 3000  
 20 TTTAATTGCA TAGGGTGAAA TAATATGTAG TCCATAACTT TTAAGTATT TTTCACTTAC 3060  
 ACCAAATTTA TAAGCTTGGT AGATAATTTT AGTACAATAC GTAAATTTTT TGCTGTTCAA 3120  
 ATTTAATGTA ACTAGATAAC GATGATTTGT ATTCTCATAG TTTTCTTAA CCCATTCAGC 3180  
 25 CGCTTTTTTA CCTGCACCAG GATAGCTGCA ACGATAAACT TTCATCCAAT CATTTTTGCC 3240  
 ACTGCATAA TTATATTTAA AAGATTCGAA GGATTGTGTA GTTGGTTTGT CGCCAGGCCC 3300  
 CTCAATTTGT AAAATCGTTT TATCATCAAT CGCGATACTA CAATGACCAA AAAATCscCA 3360  
 30 CATGACAGGG CCTTTTGTA CAATAATATC ACCAGGTTGT AATTGGAATT TGTCATCTTG 3420  
 AATTTCTGAA TACTTATTAT CTGCAATTGT TTTTGGTGAG TTTATTGGGG ATACGACAAC 3480  
 GAATAATATA AGTAAAATTA TCGTTCGTTT AATATAGTTC ACTTAAAAGC TCCTTGTTGA 3540  
 35 AGAAATATAT GTAAATAGTC TTAAATTAGA ATTGTAATCT TTAATAAGCT TGtAAGACTA 3600  
 AAACATATCT TAAATATTAA AGTATGAGAG TGTGAAATGT CTATTAAGAA TnAAAAACAG 3660  
 40 TCTGAAACAT CATTGAGACG TTCCAGACTG GATATAAAAT GAATTTCAAT TATAGCACA 3719

(2) INFORMATION FOR SEQ ID NO: 345:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1676 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 345:

TTGCGTTGCC GCACCAAGAT ATTGAATGCC TAGCGATTCC GAGTATGCAA ACTGAACGCA 60

CTCTTTAATA CGCGTTATCG CTTTTTGTA ATCTGCATCA TGATAACAAA TCATAACGCC 180  
 ATAGCCACCT GCTGTCGGAA GATCATCTCC CGGCTTAATT ACTAACGGGA ATTCCCAATT 240  
 5 CTTAATCTCG TTTTCGAATT GCTCAATTTT TACAACTTTT CTTTTTGGTA AAAACTTCCC 300  
 ATTTGTCCAT TCAGGTATTC TTGCTTTATT ATTTAAAGCA ACAAATAACG TTTTATCTAA 360  
 TGCATAATAT TGCTGATTCA AGATTGTTTC ATCATGAATA TATTGAAAAT AAATCTTTTT 420  
 10 ATTTTCCTTA TGTGCCAATT GTTTGATCAA GTTTTCGTAA GATTGCTGAT TGTAAATGT 480  
 ATAAATTGAG TTCGGtACTT CCTTACCAAT AACTTGAAAT AGCTGATGCA ATTTGTCTGT 540  
 15 CGCACTAGCT TCGTGAACAA TAACAGGTAA TTGATTTGCT ATTAATAACT CCCTACCAGT 600  
 TAAAAAATTA GATTGATGTT CGTCCGGTTT CAACCATGGA TTCGATATAT ACGAAGGTCT 660  
 TGACGTATAG ACAACATCTT TGTCATATAA ATCACTTAAC GTTAAGTTCTG GCTCATTACC 720  
 20 ATTATTTGTC ATTACTTCCC ATTCCCTTTC AAATGCGCAT GCTCTTCAAT AATGCTTTGA 780  
 TAAACGTCTT GATTTGTAAT TAACTCTAAC CCCATCAACG CCATTATTTT AGCGCCTTTA 840  
 ATTAATGCTT CATCACCATG TACACTCGCA GCCGCTTCTC TAAATCTATG CGTATGTCCT 900  
 25 ACTAAATTAC GTGATCCTAT TTTAATATGA GGATGTATTG TTGGCACAaC ATgaCTTACG 960  
 TTCCCTGTAT CCGTAGAGCC ATAACCAAAA TCATCATCAA TAACTGCTTC ACCAACTTCT 1020  
 TCAGCATATT TAGCAAATAA ATCATCTAAT TTCGGCGTTT TAATGAATTC ATTCACACCG 1080  
 30 TTTTGAATTC GACCAAATTC ATAATCACAA CCAGTCTGTA TCGCAGCTCC ACGTGCGATT 1140  
 TGATTTACTT TTTCTGTAA TATATCCAAT TCTTTACGCG TCATTGCTCT AGTATAAAAA 1200  
 35 CGAGCATGTG TATAGTCTGG AATAATATTA GCTGCTTTCC CGCCATCTAA AATCACACCA 1260  
 TGCACACGTT GATCTTTTTT AATATGTTGT CGTAGTTGTG CTACACCATT AAAATAACTA 1320  
 ATCATAGCGT CTAATGCATT TAACGCTTCA TCTGCATTTT CAGAGGCATG AGCACTTTTT 1380  
 40 CCGTAAATTA TAACATCTAA AACATCCACT GCCAAAGTAT CAATCGTTTT ATAAGTTTCA 1440  
 TTTCCCGGAT GAATCATTAA GGCAATGTCT ATTTGATCAA TCACACCAGC CTTGACATAA 1500  
 GAAGCTTTAG CGCTACCATT TTCCCCACCT TCTTCAGCTG GACATCCAAG AACGACTACT 1560  
 45 TTACCACCAA TTTGGTCAAT CACTTGCTTC AAACCAATTG CACCAAGAAC ACTTGCAATT 1620  
 CCAATGATAT TATGACCACA AGCATGACCC AATCCTGGCA AAGCATCGTA TTCTGC 1676

50 (2) INFORMATION FOR SEQ ID NO: 346:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1294 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 346:

5 TACAGTAGGA ATCATAAAAC CTAATACAAC AAATACAAAA CCATTTAAGG CATAACTAAA 60  
 TGTGTTCCAA ATTTGATGGT AATTCATTTG TAGTTCCGTT TGTGCTCTAA TTAAACGGTC 120  
 10 GCGTTCTAAA CCATGGATTA GACCTGCGAT TACAACCTGCA ATGATACCTG AAGCATGAAC 180  
 TTCTTCTGCT AAAAAGTATA CGACAAAAGG AGTTAATAAT TGAATAAAAG TTAAGGTATT 240  
 GTTATCTTTT AAACCTTTAT TAGCGGTAA GTCTATACGT ATTCTAACGA CAACGAATCC 300  
 15 AATAATTGCA CCAATAAGTA CACCTAGTAT TGTGAAATG ATAAATTGTT CAACAGCTTG 360  
 GAATAATGAA AAGGTACCAG TTAATAATGC AGTAACAGCA ATTTTAAATG AAATGATACC 420  
 TGCTGCATCA TTGAGTAAAG ATTCACCTTC TAAATCGTC ATAGAACCTT TAGGTAATAA 480  
 20 TTTTCCGCGT GTAATAGCAG ATACTGCTAC TGCATCAGTA GGACATAAAA TTGCTGCTAT 540  
 TGCAAAAGCG GCTGGCATTG GTAAGGCAGG CCAAATCCAA TGTATAAAAT AGCCAACACC 600  
 GACTACAGTT GCAAACACTA GTGCCATTGA CATTAAATAGT ATAGGTTTAC GATATTCTAA 660  
 25 TAATTTTGTT CGAGAGACGT GGGTACCTTC CACAAAAAGT AGTGGCGCGA TAACGGCAAA 720  
 CATAAATACT TCAGAATTGA ATTGGAAATC AACTTGTATT GGAATAATGA AAATAACGAC 780  
 30 ACCTAATGCA ATTTGAATAA AGGCAGTAGG AATTTGTGGG AATCGATTAT TGATAACCGA 840  
 ACTAATAATC ACAGCAAAAA TAAAAATTAA AAATGCTTCT AATAGTGCCA TACAATACTC 900  
 CTCAAAATTT TAATAGTTAA TATTTTATCA CTTTTAAGGC ATAATGACAT AGATATATTG 960  
 35 ATAAATGAA GTTATTTTCA AAAAACTCT AGTATCGGTT GAACTGATAC TAGAGCGAGA 1020  
 TGTTTAAATT ATTGATTGTC ATATCTGAAA TGACCGCTGT CATTTTGTCTG TTGTTTCATAC 1080  
 GCGAGCTTTT CAGCATTCGT TTTGTATTTT TtATAAAAGa AAAATAaAAA TATnAACCaG 1140  
 40 AATGGCGAAA TATAAATAGC TGCTCTTGtT TCGTCACTAA AGAATAATAA AATGAATACA 1200  
 AAGAAGAAGA ACGCTAGAAT AATGTAAGCA ATAGGCTTAC CACCAATCAA CTTAAATTTA 1260  
 45 CTGTTTTTAT GTGCCTCAGG ATGCTTTTTT AAAT 1294

## (2) INFORMATION FOR SEQ ID NO: 347:

## (i) SEQUENCE CHARACTERISTICS:

- 50 (A) LENGTH: 1935 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

	ACATGATAAT GATGACGCTA TTAAACACG TTTTTATTT TTCATTGTTA TAACCTTCTT	60
	TCGTATGATT GATATTTGTT GATATGTATC GACATGTGAA TAATATCACA AAAACAGAGA	120
5	ATATATATTT AACTATTTAT TAAATGATTT TGTTAATATT ATTAAATACT TTATCCTCTT	180
	TAAAAATAAT GTGTGTACAA AGTCATTAAT TTAGCAAATA TTTTATTTA GTAGTTAATA	240
	ACCATCGATT TGAAATTTAT ATATAATTAT TAGCTAAATA ATATCCTGCA TCTTTCTCAT	300
10	ACAATTTACT ATAAAtTagC ATATCCGATA TCAGCGTTAA TAAGATCGTT GATACTAGmC	360
	AGTTAATTTT ATAGAACGAA ATCAAATAAC ACACTACTTT CTGCATTTTA AATTATGTTT	420
15	AAGAATCAaA ATTATGTTTA nATAAATATA TATACTACTT TGAAAGGTGT GAGCTTAATG	480
	ACAACCTTTTA GTGAAAAAGA AAAAATTCAA TTACTAGCAG ATATTGTTGA ACTACAACT	540
	GAAAAATAATA ATGAAATAGA CGTTTGTAAT TATTTAACAG ATTTATTCTGA CAAGTACGAT	600
20	ATTAAATCTG AAATTTTGAA AGTTAATGAA CACCGCGCCA ATATCGTTGC AGAAATCGGT	660
	AACGGCTCAC CTATACTCGC ATTGAGTGGT CATATGGATG TTGTTGATGC AGGAAATCAA	720
	GATAATTGGT CATATCCCCC TTTTCAACTG ACAGAAAAAG ATGGCAAATT ATATGGCCGA	780
25	GGCACTACAG ATATGAAAGG CGGTTTAATG GCTTTGGTCG TATCTCTAAT CGAATTAAAA	840
	GAACAAAATG AATTGCCTCA TGGAACGATT AGATTACTGG CTACTGCTGG CGAAGAGAAA	900
30	GAACAAGAAG GTGCCAAATT ATTAGCTGAT AAAGGCTATT TAGACGATGT CGATGGCTTA	960
	ATTATTGCTG AACCAACTGG ATCTGGAATT TATTATGCAC ATAAGGGGTC TATGTCATGT	1020
	AAAGTAACTG CAACTGGTAA AGCTGTCCAT AGCTCAGTTC CATTTATTGG TGACAATGCA	1080
35	ATTGATACAC TGCTTGAATT TTATAATCTA TTTAAAGAAA AATATTCAGA GCTTAAACAA	1140
	CAAGATACTA AACATGAATT AGATGTTGCG CCTATGTTCA AATCATTGAT TGGAAAAGAA	1200
	ATTTCTGAAG AGGATGCAAA TTATGCATCT GGTCTTACAG CTGTATGTTT GATTATAAAT	1260
40	GGCGGcAAAC AATTTAACTC TGTACCAGAT GAAGCTTCAC TTGAATTTAA CGTAAGACCA	1320
	GTTCTTGAGT ATGATAACGA CTTTATAGAA TCGTTTTTCC AAAATATCAT TAATGATGTG	1380
45	GATAGCAATA AGCTTTCACT CGATATTCCA AGCAATCACC GACCTGTAAC AAGCGATAAA	1440
	AATAGCAAAT TAATTACTAC GATTAAAGAT GTAGCTTCTA GTTATGTAGA ACAAGACGAA	1500
	ATATTTGTTT CAGCGCTTGT AGGCGCAACA GATGCCTCTA GTTTCTTAGG AGATAATAAG	1560
50	GACAATGTTG ATTTAGCCAT TTTTGGACCA GGTAATCCAT TAATGGCACA TCAAATCGAT	1620
	GAATATATTG AAAAAGATAT GTATCTGAAA TATATTGATA TTTTAAAGA GGCTTCCATT	1680
55	CAATATTTAA AAGAAAAATA AGAACGATGC TGTCAGCTGC CCTATTGCGG TGCTGGCAGT	1740

TCAAATATCA ACAAGCACAT TTTCATTGAT TAAGTGATGT AAAACTGAAA TTATTGTGCT 1860  
 GATTGTGCAT ACATATATTG ACTAATGGGC ATATAAAAAG ATAGCCTCTA ATAGTnACAT 1920  
 5 AAACTCGTAA AAnCC 1935

(2) INFORMATION FOR SEQ ID NO: 348:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1351 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

CCTTTnCCTA AACAAATTTT AGATTTAGAC AACAAACCGA TTTTAATCCA TACATTAGAA 60  
 20 AAnATTTATTT TAATTAATGA TTTTGAAAAA ATTATTATCG CGACGCCACn ACAATGGATG 120  
 ACGCATACGA AAGATACACT TAGAAAATTC AAAATTTCTG ATGAAAGAAT TGAAGTCATT 180  
 CAAGGTGGTA GCGATCGTAA CGATACAATT ATGAATATCG TTAACATAT TGAATCAACA 240  
 25 AATGGTATTA ACGATGACGA TGTCATTGTG ACACATGATG CAGTTAGACC ATTTTAAACG 300  
 CATCGTATTA TTAAAGAAAA TATTCAAGCT GCTTTAGAGT ACGGTGCAGT AGATACAGTG 360  
 ATTGATGCTA TAGATACGAT TGTTACATCT AAAGATAATC AAACGATTGA TGCAATTCCA 420  
 30 GtGCGTAATG AAATGTACCA AGGTCAAACA CCTCAATCGT TTAATATTaA TTTATTAAAA 480  
 GAaAGCTATG CACAGTTGAG TGATGAGCAA AAGAGTATTT TATCTGATGC TTGTAAGATT 540  
 35 ATTGTAGAAA CAAACAAACC GGTTCGACTT GTAAAAGGTG AGTTATATAA CATTAAAGTA 600  
 ACAACACCTT ACGATTTAAA AGTAGCGAAT GCTATTATTC GAGGTGGTAT TGCCGATGAT 660  
 TAATCAAGTA TATCAATTAG TTGCACCTAG ACAATTTGAA GTTACGTATA ACAACGTAGA 720  
 40 TATTTACAGT GACTATGTCA TTGTACGTCC TTTATATATG TCAATTTGTG CTGCCGATCA 780  
 AAGATATTAT ACTGGTAGCC GTGATGAGAA TGTCTTATCT CAGAAATTGC CAATGTCTTT 840  
 AATTCATGAA GGTGTTGGTG AGGTCGTATT TGACAGTAAA GGTGTGTTTA ATAAAGGTAC 900  
 45 AAAAGTAGTT ATGGTACCGA ATACGCCGAC AGAAAAAGAC GATGTCATTG CTGAAAAC TA 960  
 TTTAAAATCG AGCTACTTCA GATCAAGTGG ACATGATGGG TTTATGCAAG ATTTTGTGTT 1020  
 50 GCTAAATCAT GATAGAGCTG TACCACTACC TGATGATATT GATTTAAGTA TTATTTcata 1080  
 TACAGAGCTT GTAACAGTAA GTTTGCATGC TATTCGTCGT TTTGAAAAGA AATCTATTTT 1140  
 AAATAaAAAT ACATTTGGTA TTTGGGGTGA TGGTAACTTA ggTTACATTA CAGCCATTTT 1200

GAGTCACTTC TCATTTGTTG ATGATGTCTT CTTTATTAAT AAAATACCTG AAGGCTTAAC 1320

ATTTGATCAT GCATTTGAGT GTGTGGGTGG T 1351

5

(2) INFORMATION FOR SEQ ID NO: 349:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 411 base pairs

(B) TYPE: nucleic acid

10

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

TCATCAAGTC TACGATAAAT TAAGTCCATA TCTAAAGGCT CGGGGTCGAC AGTTTGTA AAA 60

GTATAACCAA CTGCACAGTG GCTACAACGC ATATTACAAA GATTTGTAGT TGTA AATTCG 120

20

ATGTTACTTA AAGTTAATTG GCCATGTTCT TTAACATCGT TATATGCTTC CCATGGGTCG 180

TTTGAATAC TTATTTTAGG CTTGTTATTA CGCATTTTAT AAAC TCCTTA ATTGTTATTT 240

25

GATACCAATT TGATACCGTT TAATCAAATA TGCTCATAGC TTGATGTTTT TTATCAGTAT 300

ATAAATGAGA GTACGTTTGA ATTGTTTCTG TAATGTTAGA ATGCCTCATT AATTCCATTA 360

ATAAATACAT ATCTACACCA TTATTAATTA AaTAGCTAGC GTACGAGTGT G 411

30

(2) INFORMATION FOR SEQ ID NO: 350:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1639 base pairs

(B) TYPE: nucleic acid

35

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

40

TCATTTTCAT AGGTTATTAC GCAGATCAGC ATAATAATCC ATTCCATATG AGTcCTTATT 60

TTGGTTATGC AGCACGTCTA TTGGCAACAA GTGGCATTGA CTATACGTAT GTAAGAATGG 120

45

CAATGTACAT GGATCCACTT AAACCATATT TACCAGAATT GATGnATATG CATAAACTGA 180

TTTATCCnGC TGGCGATGGT CGTATTAATT ATATTACTAG AAATGATATT GCTAGAGGTG 240

TCATTGCTAT TATTAAAAAT CCAGATACTT GGGGCAAACG CTACTTATTA TCAGGCTACA 300

50

GTTATGATAT GAAAGAACTT GCTGCAATTT TATCTGAGGC ATCaGGCACA GAAATTAAAT 360

ATGAGCCCGT TTCATTAGAG ACATTTGCAG AAATGTATGA TGAACCTAAA GGCTTTGGTG 420

55

CATTATTGGC ATCAATGTAC GACGCAGGAG CAAGAGGACT ATTAGACCAA GAaTCCAATk 480

TTAATAATAA AGGAGCGTTA TAGTGAATAT CATCTCAACA ATTtTAATCA TATTTGTGGC 600  
 ATTAGAGTTT TTCTATATTA TGTACCTTGA AACGATTGCT ACAACTTCCA AAAAGACTAG 660  
 5 CGAGACATTT AATATAAGCG TCGATAAATT GAAAGACAAA AATATTAACC TACTTTTGAA 720  
 GAACCAAGGC GTATATAACG GTTTAATCGG AGTTTTGCTA ATATACGGTT TGTTTATCAG 780  
 CAGTAATCCA AAAGAAATAT GCGCAGCTAT TTTAGTGTAT ATCATTGGCG TTGCTATTTA 840  
 10 TGGTGGCCTT TCAAGCAATA TTAGTATCTT TTTCAAACAA GGCACATTGC CAGTATTGGC 900  
 ACTCATATCA ATGCTTTGGT AAGTATTGGT GTTTGGGGGG GTGGAGATGT AGTCGGAGGT 960  
 15 TTGGAGGATT TGAGCGAATT GTGTGTGGAC TTTAGACTCA GAGTATTTCA TCCTAATTAT 1020  
 TTCAAGCAGA GGTGACAGTA GCGTTGCCTC TGTTTCCTTA TAAAAAATT ATTTAATGAA 1080  
 GAAAACCCAT ATCTGATTTA ATTTTCAGCT GATAAATACT CCATATATTA GAATGGCTAC 1140  
 20 TTTATCTATT GCATCAATCC TTTAAAACAA AAAACCCATG ATTTGAAAT TCCCGTATGA 1200  
 TGGGGTTCCT ACTCTCATGG ATCAGTTAAA TAAATATTAT CACTATCAGT TTATTATTTT 1260  
 AATATTATTA ACAATATATG TAGTCGTAAA AGGAAAGAGG ACATGAGAAC TTCGGTGTGG 1320  
 25 ATTGGCATTa CATAACGCTT CCAAACATAT TATTTGGTAA CAATAAGAAA CTATTTACAC 1380  
 AATATATTTT GTATAGTAAA ATTATTTTAT AATATTTAAA TCCAATTGCA CAAGGAGTGA 1440  
 TTATCaTGGT ACCAGAAGAA AAAGGTTCTA TTACTTTGTC AAAAGAAGCA GCTATCATAT 1500  
 30 TTGCAATCGC AAAATTCAAA CCATTTAAGA ACAGAATTAA AAATAACCCA CAAAAACAA 1560  
 ATCCATTTCT TAAATTACAT GAAAACaAAA AATCTTAATC ACTTTtATTT ATAGcATTTT 1620  
 35 TAATCTCAGA AATGCTATA 1639

(2) INFORMATION FOR SEQ ID NO: 351:

(i) SEQUENCE CHARACTERISTICS:  
 40 (A) LENGTH: 1816 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

AAAATCGCAT ATAGTAATAT GAATAACCAG ATTGTATCTA CAAAAAGTA TATTGAAAA 60  
 50 CCAAGCGCAC CCATTAATAA TGCGAGAATA ATAATAATTT TTCTATTAAA GTGATGCGTA 120  
 TCCGAAAATC TAGCAATAAT TGAATTTACT GTAACTGGC TAATCGCTGC AGATGCTAGA 180  
 AGTAATCCAT ACTGATTTGT TGTCATACCT AAATCTTTAG TTGCAAAAAG AACAGATAT 240  
 55

TTCTyTATTT GTAAaTAACGC TGCAAACATA TCCATAACCC CGCTTCTTAG AGCCCCTTTT 360  
 AATTnATnAA TTAGGGGCTC TTATGCAGTT GGTGCATTAG CAACCAACTG TATTCCTTTG 420  
 5 TCCCCTTTTA ATTTATTaAT TAGGGGCTCT TTTGCTGTTG GTGCATTAGC AACCAACTAC 480  
 GTTCAATTTA ACCGAATAGT TTAAAATTAA ATACAAACCT TAAATTAGTC TAAAACTACG 540  
 CCTTTGGTTG TTCAACAAAG CTCGCCATGA GATTTACAAA AGAATCAACT TGTGGCAATT 600  
 10 GCAACATGCT CGGATCATAA CTCATAAATG TCGAACGAAT CAGCGGTTCA TTATCAATTT 660  
 CTACTTTTTT AAACTCAAAT TGTTCCTTGC TGATATTTTT CATCATAATT TCTGGCAAGA 720  
 15 TTGTAACACC TACACCACTA ATCAACATTT CTTTGCAAGT TGCTACTTGA TCCACTGTAA 780  
 TAGTTGCATG GTAATCTTGT TCTAAATTAT CGTTATACCA TTCTTTTATT TGATTTATAT 840  
 AAATCGGATC AGCTTGAAAC TCTATAAATG GTAACCTTGT AACATCATCT CGTCTATTTT 900  
 20 TTGGAAAAAT AAAATAATGA TCATCATTA ATAAATGTGT GTTAGCTAAA TTCATTACCT 960  
 TATTTCCACG AGTTATCATA ACATGATAAT CTCTATGATT TGCTTTAATT TGTTCAAGTTG 1020  
 AACCAACTTG CACTTGTATT TCAACATTAG GAAATTGGGC ATTATATAGG CTCAAACTT 1080  
 25 CAGGAAGTAA GGTTCGTCCA ATCAAAGAAG AACACCCGAT TGATATTGTT CCATTCACCT 1140  
 CACCAATATG TGCCTGCATT TTGTCAAAAA ATAATCGCTC TCTTTTCAAC ATGTCACGAG 1200  
 30 CATGCTCAAT AATCATTGTT CCTTCAGTTG TTGTAATCAA TTGTTTTTTT GTTCTGATAA 1260  
 AAATATCTAC TCCAAAAGCA TTTTCAATAG CTTTGTAGTCT TTGTGTAACA GCAGGTTGAG 1320  
 ATATATATAA AATTTACGCC GCTTTACGTA ACGTTTTTCG TTCGTCTAAT GTTATTAGTA 1380  
 35 AACGATAGTC TTCAATCTTC ATAATTTCCC CCCATAAATT ATTCAATTAT TGAACCTTCA 1440  
 TGGCTACAAG CATTCATGAG TTCATTACTA ACGAATAATT TCACCAATTT TATTGGTATG 1500  
 GCTGCAGCTT GAATTACTTA GTTTTTCTTT TGTGTTGGT GATTTTTAGT TTGATTATAT 1560  
 40 TGCTTAGGCT TTATTTGTTT GCTTTTTTCA ATATTAGTTT TATTTTGTGG CTTTGTATGA 1620  
 TTTTTTTGAG CCTTTGCATT AATTTTATTA AAGCAGTACA TGATTTTCTT TTGGAATCCT 1680  
 45 TTAAAATCAT TTCTAACTC TGCCATAATT TGATGTGCAA TCATATATGC TTCATGAAAT 1740  
 TGCTTTTTTG TAATTTGCTC ACTTTCTAAT GCAAACATTA AATCATCTTC ATCTACCAAC 1800  
 TCATaTcACC ACTTGG 1816

50 (2) INFORMATION FOR SEQ ID NO: 352:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9956 base pairs  
 (B) TYPE: nucleic acid  
 55 (C) STRANDEDNESS: double

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

5	GCGTTGTCGT CGATGATTAA TAAAGTATGG GTATACCATT AAGAATAACG CTACCCAAAT	60
	gAkTGCTAGT GACGTGCCGC CAATGACATC TGAAAAGTAA TGTGCATGAA AATAAAGGCG	120
	ACAAAATAAT ATGCTAAGCC ATAATATTCC CATAACCAAC GCACTCAACA CTTTGTAT	180
10	TGTCTTAGCA GCAAGTGAAA TAATAATGAT CATTAAAGGCG AAATATAATA ATGTGCTGGC	240
	GTTGGAATGT CCACTCGGAA ATGAAAAGCC TGTATCAACG GCTAAATGAT TATATGGTCT	300
	TGGACGTAAT ACAGTATCTT TAATTAATTT GTTCATGATG ACACCTGAAA CCAAATATGT	360
15	CACAAACCAA ACCGCTAAAT GCCTCTGTTT AATAAACAGT ATGATTGTGA CGATAATGGA	420
	AATCAACACG ACACCTTGA CATCTCCAAT TTCCGCACTA AACGTCATAT AGTAATTAAA	480
	CAAATTGTGA ACATACTGAC GTTGTGGCTC ACCGAAATAA TCTGTAAACC ATGTTAATGA	540
20	TCCCATATCT ATATTTTAA GCCATTCTTG ATTTGTCACT AACTGTAAA ACATACCTAT	600
	AAATACAATC AGCGCGATTA AAAATAAAGG CACTGTCATT TTCGGTGATG TTAATTTTTT	660
25	ATCTATCATC TTACAATCTC CTCGTATCAT CATTTTCATT TTACAAATGT TATCCATAAT	720
	ATCAATGTGC CACAAATTC ACTTTACCGA CAATATCAAA ATTATAAAGT TCATATTGTT	780
	ATGTATATTG CAAATAAAAC ATTGTATAAT TGAAATAACA ATATTTTGCT ATTTTCAATT	840
30	TAGTACGATT TATATTTATT ATACAGAGGG GGTAAGGCGT ATCAATAGAG TTATTTTGT	900
	CTATATAGCG TTAATCATT CATTAGTTAG ATTCTTTACC CCTATTCATC CATCATTTTC	960
	AAATTTGATT TACTGGATAT TTGTATTATA TTTTATTCCT ATTATACTAT GCGTTATCGG	1020
35	TTTCAAGGCC GAAAACTTA TTGCAACAAT GGTCATTATA CCTAATTTT TAGGAATACT	1080
	TTATCGATTA TATGCCTACG TCACACATAT ACTCTTTATG TAAAAGGATA GTGGGCATGT	1140
40	CTCGCAACAA ATATGCTTGC GTCGACTTGT CACCATTTCG AAATTTTATG ACATACGCCT	1200
	TTCACGGGCT ATATTCAGAC CCACGCATTC ATCCACGTAA TAAACACATC ATGTAATAGA	1260
	AAAACAGCAC ACCCAAATAT ATGGCGTTGC GCTGTTTAA CAAGCATACT TCTATAGCTT	1320
45	TAATAAGCCA GCAGAAGCAT ACCTAACCTT CTAAATATG CTTTCCAAA TTATCCTCAA	1380
	GTTTGAATAC GATAATACGT TCACCTGTAA CTGTACTTAA ATCACTATGG AAGCTCATCA	1440
	CTTTGATACC TGTAATTTTA AAAATGATAT CATTCAAATC TTGCTCACCG GATTCAACTA	1500
50	ATTCAGAACG TGTTGTTTA ATATTTAATA ATCCTTCATT CGTACTACAT ACACGATATT	1560
	CAGCTGGCGT TAAGATACCT TGTAAGCTAA TAATCACCAT ATCTCTTAAA ATGTCTGATT	1620

55

	TTTCAGCTTC GATTTACCT TTCGTTCTTT TCATATCACT AACTCCAATA ATATTAAAAT	1740
	TGATTACTTC ATCTTTGTAT CGTTATCAG ACATCAACTA TTACATTAAAG TTTATCATTT	1800
5	TTAGTATATT TTAAGAAGCT AGAACATTGT AGATATGATG ATATATTAGT TACTTAGCAT	1860
	CGCAACATAT CATCGTTAAA TCCAACTTTT AAAACGCCCT TCCTCATTAA CGCTCATTAA	1920
	ACGCAGCCAA TGATTAGACA CCTTCCTAGC GAAATGCTCA TTATTCGCGA GTAGTCTTGC	1980
10	TACAACATAG TCGGGTGCCT GAATAACGAC AAGTAAACGA ATTGGCGAAT GATACATCGT	2040
	CCGATCAGCA GCCATAACAG ATTGCCATGA TAAGCCATAC ATCAGATCAC TCGCATTACC	2100
15	TTGCATGACA CCAACACCTG ACGTGACGGT TTGTGTCGCT TTATTCCCAC TTCCGTAAAA	2160
	ATGCGGCGCA ACTGTCGACG CATAATATTG TAAATTAATC CATTGTGCCA CAAGTGCCGG	2220
	ACCAGAAATG ATGGTATTTA ATAATGTGCC ATCTTTATCT TTACGCCAAT CATAATTGTG	2280
20	TAAAAATGTC CGCCCTTCTA AATCAATGCC TTTTGTTAAT TGGCGTCGTC CAATTATAAA	2340
	TGATGCATTT TTAGCCAATC CCCATTCTGG ACGTACCTCA CTCCAATCAC TCGCAAACCG	2400
	CTGCGCTTCT TCCACTGGAT GATTACACG ACCAATCGTT GGCAGTTTGT CCAAACGTTT	2460
25	GCGATTTCGcg TGtTCAGAAA TCATCGGCAT CGCGTCATT C AATGATTCAT ATGCATCTAA	2520
	AGCAATAGAA GATAATGTGT CTGGCACATA TACCCATGCC AACGTATCAG TAGACGTATG	2580
30	ATGTTCTGCT ACCGCAAAAA CAGTTGTCTC TGGAATATAC ACACCTGATT GTTTTAATCC	2640
	TTGTCTGACA TTTGGACGAT TACATATCAT CGCTAATAAC TTAGCATTA AACCCTTGA	2700
	TGCGCCACCA CAAGCCCCAC ATTCAAGTGA TGCATGATGT GGATTATTGT GAGAATGACT	2760
35	AGCATGACCT GCTAACACAA CGAACGGCGC AAATGCTTCG GTTAAATCCA TCAATTTCAA	2820
	CGCTTGTAAC GCGAAATCAA TTTGCTCTTG CTCAGTAAAT CCAACAGGTA AGTCTGATGT	2880
	TCGGTCAAAC TCACGATCAA TCGTCAACTT TGTTTCAGGC TTTTTC AACC ACTTTTGT	2940
40	TATTTTTTGT AAAGACGCGC GACTTTTTCT AGGCATAATC GAATTGACAA TGGTACTTAA	3000
	GCTTAAAAAT GGCCCACTTA ATTCAGGCAA TAACAGACTA GGCATGACAT TATTTTTCAT	3060
	CAATTTAAAT GTGTAAAACA TCGATGACAT TGTCTGTTGC TGTTGTCGAT AAACATTCAT	3120
45	ATCGTAGCGG TCTGCAAATT CTTTAATGCG ATATGCCGGC GGTACCATGA CAGGTAATGA	3180
	ATCATGTTTG AATTGTTTCGT CTACGGCATC TTTTGAATA GGTAATCCAA AGAAGCCTGC	3240
50	AATACCAATC GTTTCAAAGG GCCCTGCTGC TTCGATATGT CTACGAAATG GTTCTGAACG	3300
	AACATCTATA CAAAATGCAA TTTGCGCTTT CGTTGATGTG CCCATCTGAT TTAGCTCGCT	3360
55	ATTATTTTCA TCAACTGCTT GTGTGTCATT TAACAATACT GAATGTGGCT GATTAGCGTT	3420



	TGCTTTAATT TTTTGTTTTA ACTGAGATTC GTATGTCATT TCCCAGGCAA TTAGCCATAA	3540
	ATTTTAAAT ACATTTTAT TCATAGTTGC TGCAAAATGA ATAAACGTTT GAATTTTCATT	3600
5	GACGTCATGT TGTAGTAATA CATCGCTAGG CATATCACTG TAGTAACACC ATGATGCAAC	3660
	AGTTTGCTTA AACCAATTTT CCGATCTACT TTCACAATCT TTAGCGACTG ACTTAACTC	3720
	ATCACCAACT AGCAATTGTT CGACAACTAA CCGAATTGCC AAATAATCCG TTAACAAATG	3780
10	TTGTTCAAAG TGATGCTGTT GTGAACGGTA ATACAACATA CCTGCCCAAC CCGGTAACGC	3840
	CAAAAGATGT CCTTCAACAT AAGCTTGGTA GTCTTCCTGA TCTATTGAAA AATGAGTTAA	3900
15	TACTGACTCT ATCGTCATTT CAGGATCATT GGGTAAGCCT TTAATCACTT GGCGCTGTGC	3960
	TTTAGTAAAA CTATGGTCAT GTTGCGCTAA ATGCAACCAT GCATGGTAAA AACTTTGCTC	4020
	ACGCTTCGGC ATTGTCCAAC TCGATAGAAA TTGATCGATA TAAAGTTTCG TCCATTTAAT	4080
20	CATTTGACGA TTCACTTGT CGCTAAGTGG CTCACCTTGT TCATCTATTA TTGCATCACT	4140
	CATCGGACGT ACATCATAGT GATGATATGA TTCAGCCATA TCACGTTTGT ATTTTCTAA	4200
	TAGTAGATCA GCAACAACAT CAACATTTGA ATGATTGATA TATGATGCAG GTACGCTTTT	4260
25	TAATGTTTAA ATGTTATCAA TATAAAGATT GATGTAGTGT TGCGGGATAT TGTAGTGATG	4320
	TTCAAGTAAC ATATCAGTAA CAAGTTGATT AAAGACACTT TCATCTAATT CACCACGTGC	4380
	CACAGCGCTT TCTATTAAATG CTTTATTTGG GAAAATATCC ACATCTCGAA CATCACGTAA	4440
30	CCATTTTGGC ACATCTTCAA ACGTATCCGC TTCTAATCCT TCCCATGGAT TTCGTGCTGC	4500
	AAAAATCGAA ATTGGTGATA ATGGTGTAAT AACACGTTTC GCATTTTCAA TGAAGTAATT	4560
35	GATATTTAAT TGTGTTGTCA TACCTTTCAC CTCCTATAAA TACTTCTTCA AATAATTCGG	4620
	ATGACTTTCT ATCGCTTTCG AGCGTGCTTC ACCTAGATTA ACTAACCACA CGTACAATAC	4680
	CGCAAAAGCC TTAGAGTATC GATGCCGCGC CACCCAAATA CTTAATAAAC TGCCAAAGAT	4740
40	TAAATAACA AACTAATGA TGACACTCAC TGTAGGCGGC GTGTGCGCAT GTGTTGTTAT	4800
	ATTTTGTAAT ACAGCGTAAA AATAATTATG TGTGATGACG TAGATAAATG TCACGATTGC	4860
	AATCAAAATC ATACCAACAA GACGTGCCAT GCGTCCTTTA CTAAAGGCTA CCATTTGATT	4920
45	CCAAGATACA AGTAATGACC ATCCTAGAAT GAGTGCACTT AACACTTCAT ATGCACTTCT	4980
	GTCACTACTC ATCCAAAATA GAAATGCCAC GATAATAGCT AATACAGTC CCATGACAAT	5040
50	CCAGCCATAA GCGTCTTTAG CAGATGCTTG TTTTGAATA TTGAATCGCT TCACGATAGA	5100
	ACCTGATTGT AAAAATAATG TTGCTTTAAA AATACCGTGC AATATTAAAT GAATAATCGC	5160
	TGCTGAATAT ACACCCAATG CACATTGAAC TAACATAAAG CCCATTTGAC TCATCGTAGA	5220
55		

	AGAAATACTA GAAAGGATAA GTAATAATGA TAACGCAAAT CCATTATCAA ATATCGGCGC	5340
	AAAACGAGTT AGAATAACAC CACCTGCATT CACAATTCCT GCATGCATAA TTGCCGATAC	5400
5	TGGCGTTGGT GCCGTTACAG ATTCAATCAA CCATCGATGA AAAGGAAATT GTGCTGCCGG	5460
	TATCATGACA GCTAATACAA GTAGTACATT CGTCAACAAT GACCATGTCTG GATGAACTAT	5520
	ATGTTGTGGT ACCCGCCACT CGCCAGTCGC AATATAAATA GTTACAATTG CTCCAACGAA	5580
10	TGCAAGCCAA CCACATAAAA ATGTCATGCT TGATAATTTT GCAGACTCAC GTGGCACTTT	5640
	CCAAAAACGA TTAACGTTCA TCAGCAATGT TAAACATAAT AATGTAATAC CCCAGCAGAG	5700
15	TGCCATCAGT CTTAAGTCTT CAGACATCCA TGCTAAAGAT GCAAACGACG TAATCGCAGT	5760
	GAACAATGGA AAGTAATGTC TATAATGATG ATCACCTAGT AAATATCGCA TTGAAAACCTT	5820
	TTGAATAATA AAGCCAAGCG CCATTACAAA GCCAGCTAAT AACCAAGATA AACGATCTAT	5880
20	TTTAAATGGA CCTAAGACAT GTTGACCATG AATACCGAAA AAGCCAATGA CTGCAAATAA	5940
	TACTGGCATG ACTAGTATGT ATAAATGTAA TTTAATATAT CTCATTGGCA TAACTGGTGC	6000
	TAAAAACAAC AAGCCACTTA TCAATGCAAT GATAAGCGCA ATAACAAACA GTGAAAATAG	6060
25	CAATTGAAAA CTTAACACTG CATAACCTCC TTATTTCTAA TCTCTCGCAT AATTGCTTAT	6120
	GTATAAAAAT AAAAACCTAC AATAGTAGAT TCTGTACATA ATGGCAGAAA ATTTACTATT	6180
30	GCAGGTTTCA GTTAACTAG ACACTGCATC ACGGTACGTT GATATACCTT GTTGCAGTGT	6240
	TCTCTTTAAG CGTGCTCCCA TGCACATATG TATATAAAT GTTACTTCTG TCTGTTCAAT	6300
	TCATCTTCAT AAATATGCTT TGCCTAGACG AGACCTAACG TGTATTTCGT TTTAACTTA	6360
35	TAACATAAAA TATAATTAAA TTTCTGCTTC ATGTCAAATT CATGAGCTTA ACCTCTATTA	6420
	AACCAATGAT TGTAAGATT TTGTAAATGC ACCTGTACAG TTAGGCAGTA TTTCCCGTCC	6480
	TTTTAAATA AAAAATTCGC AGTTATGATC ATAACAATTC AAGTTAGGAA AAAAATCAAT	6540
40	TACGCACAAG ATAACATATG ACAATGAAGT TAACTCATAA GCAAAGGAGG TAATCTTAAT	6600
	GGGTATCATC GCTGGCATCA TTAAAGTTAT CAAAAGCTTA ATCGAACAAT TCACTGGTAA	6660
45	ATAAGATTTT ATAACAAACA AAGGAGGTCT TTCACATGGG TATCATTGCA GGAATCATT	6720
	AATTCATTAA AGGATTAATT GAGAAATTCA CTGGTAAGTA AGTTATAAAA ATCTCATAGA	6780
	TATGAACATC TTATTTGAAG GGGGCCATTC ACATGGAATT CGTAGCAAAA TTATTCAAAT	6840
50	TCTTTAAAGA TTTACTTGGT AAATTTTGTAG GTAACAATA ATCTCAAACA TTAACGATCA	6900
	ACAACTCATC ACTATGTTAA ATCAACATAC AGGAGGACAA AACGATGGCT ATTGTAGGTA	6960
55	CTATCATTA AATCATCAAA GCAATTATCG ACATTTTCGC AAAATAATTT AAGCGAATTG	7020

	TTATTGATGT GAGGTGAGTC TTGTTAGTTT GTTGCAAATA AATGGTCTTG GTGTTTTTTG	7140
	TATAGGACGT TCTTAGTGGG ACATACGGAA TATTCGTGAT CTTTGTAGTC TGACGCGTTA	7200
5	TATTTTTGTG GCGTGTTTTA TGTTCGATAC TCGAGTTCTG AGACATTCAT GATTTGGCAT	7260
	GCGAAATCTT AATGATTTTC ATGATCTAGC GCAAGATATA TTGGCCACGT GCGGAATTGC	7320
10	GTTGCACGTT TAGACTGAAA CACTCGTGTG ACCGTAAGTG TTAATAGTAC ATTGATAGCT	7380
	GCATTTACTT CACTCATTTT TATGACTGTT AAACAATGAT TGTACCTTCA ATTAACAGTT	7440
	GGTACGATGG TTTTGCCATT TTTCATCAAC GTAAATATAA AAAGGACTAA GACACATACA	7500
15	TGTCCTAGCC CTATGGATAA AATGCAAAT TCTGCTTTAT CAAAACATC AACTTTTAGA	7560
	TAGATTGAAA ACAAAAAGAT CCTAAGAACA CCTTAACCTT TTATTAATTG TCATAAATTG	7620
	CAACAATTA AGCCACAATT CAAAAATGAT TATACTTCAT TCAACTTATC GTGCTGGTCT	7680
20	AATTTGCCAT TGATATGGAT CTTCAAATTG TTGCCAATCT GCATCAATTT CTTGCGCATT	7740
	GACTAAGCAT GCGTCGAGTT CTTTTGTAA TTTTCTTCA TCTAATTCTG TACCAATAAT	7800
	GACAAATTGT GTATGACGAT CGCCATATTC TGGATCCCAT TCAGCTGCGA CATCTTGACG	7860
25	TTCTGCTAAT ATTTGTGTTT GTTGCCTTC AGACATACTA GCCACCCAAT ATGTAAGTGG	7920
	ATGAATATTG CAAGATGACC CTGCTTGAGA TAATAAACAT GCTACGTGAT TGTATTGTGC	7980
30	TAGCCATACG ATACCTTTTG ATCGAACGAC ATTATTTGGC ATGCTTTCTA ACCAATCATT	8040
	GAACCTTTTA GCATGGAAAG GTAGACGACG TTTATATACA AACGATGATA TACCATATTC	8100
	TTCTGTTTCA GGTGTATGCG ATGCATGCCC ACCAGACTCA AGTTCTTTGA TCCATCCTGC	8160
35	TGACTCGCTC GCTTTTTCAA AATCAAAACG CTGCGTATTC AAGACTTCTT TTAAATCTAC	8220
	TTCAGAATTT GTTGTCTTAA TAATTTTAGC AGTCGGTTGC AATGCGCTTA ACATTTTTTC	8280
	TAACTTCGCT AGTTCTTCTT CACTAATTAA ATCAATTTTA TTAATAATCA ATACATCACA	8340
40	AAATTCAACT TGGTCAATTA ATAAATCAGC AATCGAACGC TCATCTGTTT CGTCAACGCT	8400
	TTGATCACGA TCCATCAATA AATCTTCTGA GTTGATGTCA TGTACGAAGC GGTTAGCATC	8460
45	CACAACTGTA ACCATTGTAT CTAAACGGCA AATCGCTGTA AGATCAATGC CAAGTTCATC	8520
	ATCAATATAT GAGAAAGTTT GTGCAACAGG TACTGGCTCT GAAATCCCTG TTGACTCAAT	8580
	AACAATTTGA TCGATGCCAC CTTTTTTCAC TAAACGCTCA ACTTCTTTTA ATAAATCGTC	8640
50	TCTAAGTGTA CAACAGATAC AACCATTAGA AAGTTCGACT AATTTTTTCAT CTGTACGCGA	8700
	TAGTCCCCCA CCATCTGCGA CAAGATCTTT ATCGATATTT ACTTCACTCA TATCATTTAC	8760
55	AATTACCGCG ATACGTCGAC CTTCTCGATT TTGTAAAATA TGATTTAACA ACGTTGTCTT	8820

ACTTCAATTT ATTTGTAAAT AGGAATAATT CTGTTTTACA TTATATAGGA GCGTTTCCTC 8940  
 TTTCGCAATC TTCGATAATA AAAAAATAGT ATACTTAATT AAATTATTGA GCGCTTTACT 9000  
 5 TTATAATGGA GACAAAGATA TATCTCACGA AAGAGAATCG AGGTGTATAA ACATGTTATT 9060  
 TGTCAATTTA GTTTTATATG TTACTGGTAT TGCATTTATT CTAATCAGTG TTTTGGTTTC 9120  
 AAAGACTGAA GGATTATCTA CGAAACATAC TTTATATACC ATTGGCAGTG CTATTATAAC 9180  
 10 GATTGCTATT TTCATTTCAA TTGGCTATGC CATTCAATAC TTAATGCAG CGCTTTATGG 9240  
 TTTGTAAGGT GAAGGTGATG AGTAACGGGT AGTTCGGGAG AGGTAACTT GCGTTGATTT 9300  
 15 TGATAAAGTG ATCATAGCTT TTAGTACTTG AGGATTTTTA TTGTTGCTGT TACGAATGTG 9360  
 GTCATGTTTA ATGCGGGACA GTAATTTAAG TTGTTTTTTT ACAATTGAGA GTGTGATATT 9420  
 TCGATTCGGT TCGAATTACT TTACATGGGA ATAATATAAA TAAAAAGAA GCGGCCTAGT 9480  
 20 GTCAGTTGTG AATATACTGA ACATTGGTCG CTTTATTTAG TAGTATGATA TGTAGTTTAG 9540  
 CTATTAATTT TTTTCAGGTC ATCCTTAATG CTGTCTATCT CAGACATGGC ACTTTTAACC 9600  
 CAATCTCCTT GAGCTGCACC TTTAAAATTA GCTTTAAAAG ctTCGCAATG TTGCGCCATT 9660  
 25 TGTTCAAATTA ATACTTTTTC TTCACCTTTT AATCCGTTTT CAATATCTTT GTATTTATGC 9720  
 TTATGTTTCA GTGCAATAAC TGTGCGAATA TTTTCTTTT GCGCTTCCAT TTTAGATATG 9780  
 30 AGATTAAGTG TTTCTACTGT AGTACTTATA TCTGGCATTC TTAAGGTCAT ATCTGGTTCT 9840  
 ATTAGAGTCA TTTAATCTCC TCCAAATTAT CAGTCACTTA GCTTATCTAA CTGCTTTTCA 9900  
 TAAGACTTTT TTAAGTCTTC TTTATATTCT TCTAATTTC CATTCTTGCT TTCTGA 9956  
 35 (2) INFORMATION FOR SEQ ID NO: 353:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2411 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 353:  
 TTTCTTTTAA CAGaTTTATC cCATTAAATG TTCTACAAGC CCACGaTGAG CAATATCATT 60  
 TTTAGCAACC ATTAATAAAC CAGAAGTATC CATATCTATA CCGTGAACAA TACCTGGACG 120  
 50 AATTTCTCCA TTAATACCTG ACAAATTTTT AATTTGATAC ATTAAACCAT TAACTAATGT 180  
 ATTGGTATAA TGCCCTGGTG ATGGATGAAC TACCATGCCT TTCGGTTTAT ATACAACTGC 240  
 AACATCGTCA TCTTCATAAT AAATATCTAA ATTTAAATTT TCAGGTAGAA TATCAGCTTC 300  
 55

	AACAACCTTA	TCGTTTGCAA	CGACTAAACC	TGCTTTAATC	CAATCTTGTA	TCTGGTTACG	420
	AGACCAATCA	TTATTTAATT	CAGGCAGCAA	CTTATCTACA	CGCATACCTG	TTTGTTCCTT	480
5	ATCTGTAATG	TTAAATTCAT	AAGTCTCCAT	TACTTAACCT	CCTTCTCCTT	TTTATTGGAA	540
	GTATCCTTTA	ATAAGGCAAT	AATAATTAAT	ATTACACCAA	TTGTTAAACT	TGAATCTGCG	600
	ATATTAAATA	TTGGAAAATC	ATAACCAAAA	ATATTTGTAT	CAATAAAGTC	AACAACCTCT	660
10	CCTGTTAAAA	TTCTATCAAT	AAAGTTTCCA	AGTGCACCTG	CAAAAAGTAA	ACTAATAGCA	720
	ACTTGCAATA	ACAAATTATA	TTGAGCATCT	TTAATAAAGA	AATATACTAA	GGCTATTAAAT	780
15	ATAATAATGG	TAATAATAAA	GAAAAATGTC	ATTTTTCCAC	TCAATATTCC	CCATGCAGCA	840
	CCATTATTTT	GATGTGATGT	TATGTTTTAA	AAGTGCAGTA	TCACTTCAAA	TGAATCTCCA	900
	ATTTTCATTG	TAGTAGCTAT	AATATATTTA	GTAACCTGGT	CAAATATAAC	GACAAATACT	960
20	GCTATTAAAA	TGGAAGTGCC	AATAAAATAT	TTTTTGTCGA	TTTCGTTTCC	TCCAATCAAT	1020
	CGTCCATGAG	ACAACTCTTT	ATATTATAGC	TTACACCTGC	TAATAAAAAA	AGTAAGCATA	1080
	TTACATTAAA	TCTAATGTTA	CTAACTCAAT	ACTTGATAAA	CTACTATGTT	TTGACATTAA	1140
25	ATATGAACTT	AATTATTCAT	TTATCATATT	TAAGATGACA	TTAAAAATTA	GGAAAGCAGG	1200
	CTGGAACATA	AATCCCTAAA	AAGACAGTAG	TAAGATATTT	TCTAATTAAA	AATTATCTTA	1260
30	CTGCTGTTCT	CTATTTATAC	AATACTTCGT	ATTGAATGGC	TTGCTATGCT	CCATCTGGCA	1320
	CATTACTGTA	AAATTCTATA	AATAGAATTT	TTGATGATGG	GTCCCTTCCT	AGGGTGCCGT	1380
	CTCAGCCTCG	GCTTCGACT	GGCACTGCTC	CCTCAGGAGT	CTCGCCATTA	ATACTACGTA	1440
35	TTAACATGTA	ATTTTACTTT	TAAATACTTT	AAAAAAATAA	GACATGAATC	GTCTACACTT	1500
	AATTGGACAA	ATTCTATGAG	AATAGATATT	GTTAATTTAA	GAAAGTAGGC	TATTTTGAGT	1560
	TtCACTCGAA	TGTCAGTTCG	AGGAATAAAT	AAAGTTAAAC	GAGAGCTAGG	TTTTGTATTA	1620
40	ATGGCAATTA	ATATAAGGAA	AATAGCAGCT	CAACGAGCTG	TACATTATAA	AATACATATC	1680
	AAAAAAGCTG	ATTCTATCA	AATAATTAAT	AGAAATCAGC	TTTTTTACAT	TGCCTAAGAA	1740
45	CTTAATGTCC	CAAGCCCTAA	AACCTGTTGT	TATTTATTTG	ATTTAGCAGC	GATACGTTTA	1800
	TATCTTAAGT	ACATAAATGC	TAAAAGTATA	AACCAAATCG	GAATAAAATA	AATTGCACGT	1860
	CTTGTATCAA	CATTAATAAA	TAATAACCCG	AACACAAAAA	TGAAGAATAC	AAATATTAAG	1920
50	TAGCCCATAT	ATTTGCCACC	TAATAGTTTG	TACGTAGCAT	TTTTATGTAG	ATCTGGGTTT	1980
	TTACGACTAT	AATTGATATA	TGCAATGATA	ATCAGACCCC	ATACAACTAA	AAATAACACT	2040
55	GTAGAGATGG	TAGTCACATA	CGTAAATACT	TTTGTGCGAT	CTGGGAAAT	ATAGTTTAGT	2100

TTATTCGTCT TAGAAAAGTT CGGAGGTGCT TGTYGTTGAC TTGATAAACC GaAAAAGCATA 2220  
 CGGCTATTTG AGAATATACC ACTGTTACAT GATGAAGCAG CAGCGGTTAA TACTACAAAA 2280  
 5 TTAATCAAGC CCGCAGCAAA CGGAATTCCG ATCAATGCGA ATrATTTTnC GAATGGACTG 2340  
 TTATCAGGAT CAACTTGCTG CCAAGGGGTA ATAGACATGA TAACCGCTAA CGCCCCAACG 2400  
 TnATATTA A 2411

10 (2) INFORMATION FOR SEQ ID NO: 354:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 605 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

GnGAATTATT TTTAATAATG AAAGGATTAC TnCATGGGT TTTTACTAGG AnTACCCAGA 60  
 AGGTCAAAAT ATTTTIGATG CGCTAAGTCA ATATGAAGTT AAGCGACGCG GCGATATGGA 120  
 25 AGAGGATCCA TCATATAAAC AACTCATTTT TATTGTTTA CTTGAAAATG AGCATGGCGA 180  
 GATATTAGTG TATGAACGAT TATCTGGCGG TGGAGAAGCT CGATTGCATG GACAATCTTC 240  
 AATAGGTGTA GCGGGTCATA TGAATGATGT TCCAGGAGCA GAATCTATTA ACGAAGTATT 300  
 30 GAGAGTTAAT GCACAGAGAG AATTAGAAGA AGAAGTAGGT TTAAGTGAGC AAGATTCACA 360  
 AAATATGGAA TATATCGGTT TTATTAATGA CGATAATAAT GAAGTGGGCA AGGTACATAT 420  
 35 TGGTGTGTGTA TTTAAATCA CTGTAAGTAC GAATGATGTA GAAGCTAAAG AAACAGATAC 480  
 TTTACGAATA AAATGGGTTG AAAAAGGCAA CATAGAGTCA TATGATGATT TCGAAACGTG 540  
 GAGTGCATTA ATCCTTCAAG ATTTATAATC AAACGAGGTG ACATATATGT CAGATATTAT 600  
 40 TCCAG 605

(2) INFORMATION FOR SEQ ID NO: 355:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 668 base pairs  
 45 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

TTTATTAGCA CATCCAACT ATTCATATGT TGGACAATTT TTAAACGAAC TAGGATTTAA 60  
 55

ACAATTAGAC ACTGAACATT TAGCTGATTT AAATCCAGAG CGTATGATCA TTATGACAGA 180  
 TCATGCTAAA AAAGATTCTG CTGAATTCAA GAAGTTACAA GAAGATGCAA CATGGAAAAA 240  
 5 GTTGAATGCA GTTAAAAATA ATCGCGTGGA TATTGTTGAC CGTGATGTTT GGGCAAGATC 300  
 TCGTGGCTTA ATTTCTTCTG AAGAAATGGC TAAAGAACTT GTTGAATTAT CAAAAAAGA 360  
 AAAAAAGTAA GGTGGAAGTA AATGGCTATA AAAGAAATAA GTAGCCAATC TGCCATAGAT 420  
 10 CATAAAGAA AAAGACGCAC AACACTCAGC TATATAGTGA GTTTGTGCTT TCTTTTATT 480  
 TGTATATATT TAAATATGGC GATTGGTCTT TCGAAAATTA ATTTTAGCGA TATCATTCAC 540  
 15 TATGTTACTG GTCATACAGA TACGAAAGCA ACGTTTTTAT TGCATAATGT ACGTATGCCA 600  
 AGGATGATTG CAGGGTTATT TATTGGCGGT GCATTAGCGG TATCTGGTTT GTTAATGCAA 660  
 GCAATGAC 668

20 (2) INFORMATION FOR SEQ ID NO: 356:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 787 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

ATACAAAAA ACATATCGAA AATAAAGCTA AAAGAACTA TCAAGTTCCA TATTCAATTA 60  
 ATTTAAATGG TACATCTACA AACATTTTAT CGAATCTTTC ATTTTCAAAT AAACCTTGGA 120  
 35 CAAATTACAA AAATTTAACT AGTCAAATAA AATCAGTACT GAAGCATGAT AGAGGTATTA 180  
 GTGAACAAGA TTTAAATAT GCTAAGAAAG CTTATTATAC TGTTTATTTT AAAAATGGTG 240  
 GTAAAAGAAT CTTACAGTTG AATTCAAAAA ATTACACAGC AAACCTTAGTT CATGCGAAAG 300  
 40 ATGTTAAGAG AATTGAAATT ACTGTAAAA CAGGAACTAA AGCGAAAGCA GACAGATATG 360  
 TACCATACAC AATTGCAGTA AATGGCACAT CAACACCAAT TTTATCAAAA CTAAAAATTT 420  
 CGAATAACA ATTAATTAGT TACAAATATT TAAATGACAA AGTGAAATCT GTATTAAAAA 480  
 45 GTGAAAGAGG CATCAGTGAT CTTGACTTAA AATTGCGAA ACAAGCAAAA TATACAGTAT 540  
 ATTTCAAAAA TGGAAGAAA CAAGTAGTGA ATTTAAAATC AGACATCTTT ACACCTAATT 600  
 TATTTAGTGC CAAAGATATT AAAAAGATTG ATATTGATGT AAAACAATAC ACTAAATCAA 660  
 50 AAAAAAATAA ATAAATCTAA TAATGTGAAA TTCCAGTAA CAATAAATAA ATTTGAAAC 720  
 ATAGTTTCAA ATGAATTTGT GTTCTATAAT GCAAGCAAAA TTACmATTAA TGaTTTAAGT 780

55

## (2) INFORMATION FOR SEQ ID NO: 357:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 534 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

```

AAAGTAAAAA TAAATCTCCC TTTTAACTT TCGTTTCTGC CATAGCCATT GCTTCTTCTG      60
TGATAGTTGC TACAATATCT TTTCTTTCAC GGTTAAAAATG TTCAACTTGT TCTGCTAAAA      120
ATGCAGCTTC TTCTTCGACG TCAGTCATCA ACAATTCGCa AGCTAATGAT GCGTCATCTA      180
AACGACCTAC AGCATTAAAGT CTAGGTCCAA TAATAAAACC AATTGTTTCT TCATCAATAT      240
TGTCAATTGTA TCCCGCTTCT TTTAGCAATG CTTTAACAGA GGTCGGACAT TGATCATTTA      300
AGACTTTTAA TCCTTGTtTC ACTAATGATC GATTTTCATC AGTTAAGGAT ACTAAATCCG      360
CAATGGTACC TATCGCAACT AATGCTTTAA AATAATCAGG TACATTTtCA ATCAATGCTT      420
GTGCTAATTT GTATGCAACA CCTGCACCAC ACAATTGTTG GAACGGATAA TTAAACGATG      480
GATGCATTGG ATGTACGATT GCATATGCTT CTGGTAATGT ACTACCAATT TCAT          534

```

## (2) INFORMATION FOR SEQ ID NO: 358:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3621 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

```

GGTGAGTCAA ATTAAATGAA TCTAATAAGT CATAACTATC TATTTGTAAT GTGCAACGCT      60
TAACGCATAT ACAAATGAA TGTGCTGATA ATGATTTACT CAAATTAAAA GGTGATTTTT      120
ATTCAATGAT GAATGAAAGT TGCCTTTTTA TTTTGGTAA AAGTTAATGC GTCAGTGAAT      180
TGTGTAAGTT TTTCAAAAAG TAAAAAGAAA TAATAAAGGT GAATTATTAG AATTCCArAA      240
ATAATTCATT ACATTCATAA AGCATTTTAC AAATGGTAAG AAAATGAGTG TTACAAATCT      300
AAATATTGCA AAAGAAGCTG ATTTAGTCAC AAAAAATGTC CTATGTAATA ATTCGAGAAA      360
GATGCACTAT ATACGGTCTT CTTACTATTC AAATGTAAAA GTTGCTTATT TGCGTGGCTT      420
TTTGTTTTAT AAAAGTATAA AATTTTACTA TAATATATCT TGTAGAGAAC AATGAAATGA      480

```



	ATAGCAAAC	GTATTACTTT	GATACAAAA	TGGTTGTAAT	AAATATTTAT	CGATATGACG	600
	ACTTGAATAT	GATAAAGTGA	CATATTTATG	TATATGACTA	TTTCGCAAAA	TGTAATCGAG	660
5	GTAGAATTTC	TTGACAATTC	TGTCAGTTTA	TAAGATGTTA	TAAATATGTA	GTGTATAAGG	720
	AGGCAAACAA	GATGACTGAA	GAATTCATG	AATCAATGAT	TAACGATATT	AAAGAAGGTG	780
	ACAAAGTCAC	TGGCGAGGTA	CAACAAGTTG	AAGACAAGCA	AGTTGTTGTT	CATATCAACG	840
10	GTGGTAAATT	TAATGGGATT	ATTCCTATTA	GTCAACTATC	TACGCATCAT	ATTGATAGCC	900
	CAAGTGAAGT	TGTAAAAGAG	GGCGACGAAG	TTGAAGCATA	TGTCACTAAA	GTTGAGTTTG	960
15	ATGAAGAAAA	TGAAACTGGA	GCTTACATCT	TATCTAGAAG	ACAACTTGAA	ACTGAGAAGT	1020
	CTTATAGTTA	TTTACAAGAA	AAATTAGATA	ATAATGAAAT	CATCGAAGCG	AAAGTAACAG	1080
	AAGTAGTTAA	AGGTGGTTTG	GTTGTTGATG	TAGGACAAAG	AGGTTTGTGTT	CCGGCTTCAC	1140
20	TAATTTCAAC	AGACTTCATT	GAGGATTTCT	CTGTGTTTGA	TGGACAAACA	ATTCGTATTA	1200
	AAGTTGAAGA	ATTGGATCCT	GAAAATAATA	GAGTCATTTT	AAGCCGTAAA	GCAGTTGAAC	1260
	AAGAAGAAAA	CGATGCTAAA	AAAGATCAAT	TATTACAATC	TTTAAATGAA	GGCGATGTTA	1320
25	TTGATGGTAA	AGTAGCGCGT	TTAACTCAAT	TTGGTGCATT	TATAGACATT	GGCGGTGTTG	1380
	ATGGTTTAGT	GCATGTATCT	GAACTTTCTC	ACGAACATGT	TCAAACACCA	GAAGAAGTAG	1440
30	TTTCAATTGG	TCAAGATGTT	AAAGTTAAAA	TTAAATCTAT	TGATAGAGAT	ACAGAACGTA	1500
	TTTCATTATC	AATCAAAGAT	ACGTTACCAA	CACCTTTCTGA	AAATATTAAA	GGTCAATTCC	1560
	ACGAAAATGA	TGTCATTGAA	GGTGTCTAG	TAAGATTGGC	AAACTTTGGT	GCATTTGTTG	1620
35	AAATTGCACC	AGGTGTACAA	GGACTTGTA	ATATTTCTGA	AATTGCACAC	AAACACATTG	1680
	GTACGCCAGG	TGAAGTGTTA	GAACCTGGTC	AACAAGTAAA	TGTTAAAATA	TTAGGTATTG	1740
	ATGAAGAGAA	TGAAAGAGTA	TCACTATCTA	TTAAAGCAAC	ATTACCAAAC	GAAGATGTTG	1800
40	TTGAAAGTGA	TCCTTCTACG	ACTAAGGCGT	ACTTAGAAAA	CGAAGAAGAA	GATAATCCAA	1860
	CAATTGGCGA	TATGATTGGT	GATAAACTTA	AAAATCTTAA	ACTATAATTT	AATATTTAAT	1920
	AGTCAACTCC	ACATGTTTAT	GATTGcATGT	GGAGTATTTT	TATGTAACAA	AATATACTCG	1980
45	GAATGATAAC	GTGGgACAAA	TTTAACTAAG	TGTTTTAAAA	GATArAGTTT	TAAGTGctGa	2040
	cTTTTATCAT	TACAGTAATA	AACTCATTTT	GAATACACAG	TCTCATGTGA	TATTATTAAA	2100
	AAGATATaAG	AAAGAGAGGA	AGTTAGCTTA	TGACTAAACC	TATAGTAGCT	ATTGTAGGTA	2160
50	GGCCTAATGT	AGGTAAATCT	ACAATTTTTA	ATAGAATAGT	TGGAGAACGT	GTTTCGATTG	2220
	TGGAAGACAC	GCCAGGTGTA	ACACGAGATC	GTATTTATTC	TTCAGGTGAA	TGGTTAACAC	2280
55							

AAATTAGAGC GCAGGCAGAA ATCGCCATAG ATGAAGCGGA TGTTATTATT TTTATGGTTA 2400  
 ACGTGCGTGA AGGATTGACA CAAAGCGATG AAATGGTCGC TCAAATTTTA TACAAATCTA 2460  
 5 AAAAACCGGT CGTATTAGCG GTTAACAAAG TAGATAATAT GGAAATGCGT ACAGACGTGT 2520  
 ATGATTCTA TTCATTAGGA TTTGGTGAAC CGTATCCGAT ATCAGGGTCA CATGGTTTAG 2580  
 GTCTTGGTGA CTGTGTAGAT GCAGTTGTTT CTCATTTTGG TGAAGAGGAA GAAGATCCTT 2640  
 10 ATGATGAAGA TACAATTCGA CTATCCATTA TTGGACGACC AAACGTAGGT AAATCAAGTT 2700  
 TAGTAAATGC TATTTTAGGT GAAGATCGCG TTATCGTTTC TAATGTTGCA GGGACAACGA 2760  
 GAGACGCTAT TGATACAGAG TATAGTTATG ATGGACAAGA TTATGTTTTA ATCGATACTG 2820  
 15 CTGGTATGCG TAAAAAGGA AAAGTATATG AATCAACTGA GAAATATTCA GTATTAAGAG 2880  
 CTTTAAAGC GATTGAACGT TCAAATGTTG TTTTAGTGGT CATAGATGCA GAACAAGGCA 2940  
 20 TCATTGAACA AGATAAACGT GTTGCAGGAT ATGCACATGA ACAAGGTAAA GCAGTCGTGA 3000  
 TTGTCGTAAA TAAATGGGAT ACTGTGGAAA AAGATAGTAA AACGATGAAG AAATTTGAAG 3060  
 ATGAAGTACG TAAAGAATTC CaATTTTATG ATTATGCACA AATTGCTTTT GTGTCTGCTA 3120  
 25 AAGAAGCAC AAGATTACGT ACATTATTCC CTTACATCAA TGAAGCAAGT GAAAACCATA 3180  
 AAAAACGTGT TCAAAGTTCA ACTTTAAATG AAGTTGTTAC TGATGCAATT TCCATGAACC 3240  
 CTACACCAAC AGACAAAGGT AGACGTTTGA ATGTCCTTTA TGCAACACAA GTTGCTATAG 3300  
 30 AACCACCGAC ATTTGTTGTA TTTGTTAATG ATGTAGAATT AATGCATTTT TCTTATAAAC 3360  
 GCTATTTAGA GAATCAAATC CGTGCCGCTT TTGGTTTTGA AGTACACCA ATTCATATTA 3420  
 35 TAGCTCGAAA GAGAAATTAA CGATTGGGGG ATAACAATGA CTAAATTAC CGTTTTTGGT 3480  
 ATGGGAAGTT TTGGGACAGC CCTTGCCAAT GTTCTTGACG AAAATGGACA TGATGTTTTG 3540  
 ATGTGGGGTA AAAATCAAGA TGCTGTTGAT GAATTAAATA CATGTCATAC AAATAAAAAG 3600  
 40 TATTTAAAAT ACGCGAAATT A 3621

## (2) INFORMATION FOR SEQ ID NO: 359:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 643 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

CTTTCGGAAA TTAGGATCnG nGCTATCTTG GCCCAATTA CCAAGGGAAC TAnTGGCACC 60

AATGCTCTTT TCATCTCCAT GCCCTGTTGC TCATTATTAA TAACACGGTC TATTAACACA 180  
 ATGGCATTTC TTAATACGAT TCCAATTAAC ATTAGCATAC CAATTAAACT TGGTACTGAT 240  
 5 ATTGTTTCTC CTGTGATTAA TAGTGCAATA ATTACACCGA TAACTGTAAA TGGTAAAGAG 300  
 AATAAAATTG TAAATGGTGC TAGGCCACCT TTAAATGTAA TAACTAGGAT TAAATATACG 360  
 ATAATGATTG CAGCTAACAT TGCAAAGGCT AATTGTGTCA TTGCATTGTT AATATCATCT 420  
 10 GATGCACCAC CGATATTAAC CTTTACATTA TTCGGTTTAT CCAAATTATT TATTTTAGAC 480  
 ATCACTTGTC GTGTGTGCC ACCCACATCT TTATTTGTGA CTTTAGCAGA TACCGTCGTT 540  
 15 GCATAATCTC CTTGTTCTTG CGTCAATTTA CTTGGTGTGC TTGTTTTAAC TAACGTAGCG 600  
 ATATCTCCCA ATTTAATCGT ACCACCAGTC GGCTTTTTCA AAG 643

## (2) INFORMATION FOR SEQ ID NO: 360:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2524 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

30 TTTGAGGCTG TGCAAGGGCT TTTTCTTTTG CTTTAAAGAT TATGATTTAT CGTGCAAAGT 60  
 TAAGTGGTCG TATATAGTTT TAGTTTTAAA AAGGTAATTA AATAAAATAG TTTGCCGAGG 120  
 GAGATGTCAA AATGATTAAA ATACCTAGAG GGACGCAGGA TATTTTACCT GAAGATTCAA 180  
 35 AGAAATGGCG TTACATTGAA AATCAATTAG ATGAATTAAT GACATTTTAT AATTATAAAG 240  
 AAATAAGAAC ACCAATTTTT GAAAGTACAG ATCTTTTTTG AAGAGGTGTT GGTGATTCAA 300  
 CCGATGTCGT AAAAAAGAA ATGTATACAT TTAAAGATAA AGGCGATAGA AGTATTACAT 360  
 40 TAAGACCTGA GGAACAGCT GCAGTTGTGC GTTCATATAT TGAACATAAA ATGCAAGGTA 420  
 ATCCAAACCA ACCAATTAAA CTTTATTACA ATGGACCGAT GTTTAGATAT GAACGTAAGC 480  
 AAAAAGGACG CTATCGTCAA TTAAATCAAT TTGGTGTAGA AGCTATTGGT GCTGAAAATC 540  
 45 CTAGCGTAGA TGCAGAAGTA TTAGCTATGG TTATGCATAT TTATCAATCA TTTGGATTAA 600  
 AACATTTAAA GCTTGTTATT AATAGTGTAG GGGATATGGC GTCTCGAAAA GAATATAACG 660  
 AAGCGTAGT GAAACACTTT GAACAGTAA TTCATGAATT TTGTTGAGT TGTCAATCAC 720  
 50 GTTTCATAC AAATCCGATG CGAATTTTGG ATTGTAAAGT AGACCGTGAT AAAGAAGCGA 780  
 TTAAGACTGC ACCTAGAATC ACTGATTTCT TAAATGAGGA ATCTAAGGCA TATTATGAAC 840

55

GTGGATTGGA TTATTATACA CATACAGCAT TTGAATTAAT GATGGATAAC CCTAACTATG 960  
 ATGGTGcCAT TACAACGCTT TGTGGTGGTG GCCGTTATAA TGGTTTATTA GAATTGCTAG 1020  
 5 ATGGTCCAAG TGAAACAGGT ATTGGTTTTG CGCTAAGTAT AGAACGATTA TTGCTTGACAC 1080  
 TTGAAGAAGA AGGTATCGAA TTAGATATTG AAGAAAACCT AGATTTATTC ATTGTTACAA 1140  
 TGGGTGATCA AGCAGATCGA TATGCTGTGA AGCTATTAAA TCATTTGAGA CATAATGGTA 1200  
 10 TTAAAGCAGA TAAAGACTAT TTACAGCGTA AAATTAAAGG ACAAATGAAA CAAGCAGACC 1260  
 GTTTAGGTGC CAAGTTTACA ATCGTTATTG GTGATCAAGA ATTAGAAAAT AATAAAATCG 1320  
 ATGTTAAAAA TATGACAACT GGTGAATCTG AAACAATTGA ATTAGACGCA TTAGTCGAAT 1380  
 15 ATTTTAAGAA GTAGAGAGGG CGTTAAAATA TGAGTAAGAG AACAACTTAT TGTGGATTAG 1440  
 TTAAGTGGGC ATTTTtagga CAAGAAATTA CATTAAAAGG ATGGGTTAAC AATCGTCGTG 1500  
 20 ACCTTGGTGG ATtGATTtC GTTGATTtAA GAGATAGAGA AGGAATTGTA CmAGTCGTGT 1560  
 TTAATCCTGC ATTTTCAGAA GAGGCaTTGA AAATTGCTGA AACAGTACGT TCTGAATATG 1620  
 TTGTAGAAGT TCAAGGTACA GTTACGAAGC GTGACCcTGA AACAGTTAAT CCTAAAATTA 1680  
 25 AAACCTGGCCA AGTTGAAGTA CAAGTTACAA ATATTAAAGT GATTAAATAA TCTGAGACAC 1740  
 CACCATTTTC TATAAATGAA GAAAATGTTA ACGTTGATGA AAATATTCGA TTAAAATACC 1800  
 GTTATTTAGA TTTACGTCGT CAAGAGTTAG CGCAAACATT TAAAATGAGA CATCAAATTA 1860  
 30 CACGTTCTAT TCGTCAATAT TTGGATGATG AAGGGTTCTT TGACATCGAA ACACCGATAC 1920  
 TAACGAAGTC AACACCTGAG GGTGCACGTG ACTATTTAGT ACCATCTCGT GTTCATGATG 1980  
 GTGAATTTTA TGCATTACCA CAATCACCAC AATTATTTAA GCAATTATTG ATGATTAGTG 2040  
 35 GATTTGACAA ATACTACCAA ATCGTAAAT GCTTCCGTGA CGAAGATTTA CGTGCAGATC 2100  
 GTCAACCTGA ATTTACACAA GTCGATATTG AAATGAGTTT TGTAGACCAA GAAGATGTGA 2160  
 40 TGCAAATGGG TGAAGAAATG CTTAAAAAAG TTGTTAAAGA AGTTAAAGGC GTTGAAATTA 2220  
 ATGGCGCTTT CCCACGCATG ACATATAAAG AAGCGATGCG TCGCTATGGT TCTGATAAAC 2280  
 CAGATACACG TTTTGAAATG GAATTAATTG ACGTTTCTCA ATTAGGACGT GATATGGACT 2340  
 45 TTAAAGTATT TAAAGATACT GTTGAAAATG ATGGTGAAAT TAAAGCAATT GTCGCTAAAG 2400  
 GTGCAGCTGa ACAATATACT CGTAAAGaTA tGGGaTGCTT TAACAGAATT TGTAaACaTC 2460  
 50 ymTGGtGCTA AgGtTAGCGT GGGGTAAAG TTGTGGGAAG GTGGTTTTGA CAAGGTCCCA 2520  
 ATGG 2524

(2) INFORMATION FOR SEQ ID NO: 361:

55

(A) LENGTH: 1507 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 361:

10	TCGTTGAGTA AAAGTCCAGA AAATTGGATG AGTAAACTTG ATGATGGAAA ACATTTAACT	60
	GAGATTAATA TACCGGGTTC ACATGATAGT GGCTCATTCA CTTTAAAGGA TCCAGTAAAA	120
	TCAGTTTGGG CAAAGACTCA AGATAAAGAT TACCTTACCC AAATGAAGTC GGGAGTCAGG	180
15	TTTTTTGATA TTAGAGGTAG AGCAAGTGCT GATAATATGA TTTCAGTTCA TCACGGCATG	240
	GTTTATTTGC ATCATGAATT AGGAAAATTT CTCGATGATG CTAAATATTA CTTGAGTGCT	300
	TATCCAAACG AAACAATTGT GATGTCTATG AAAAAGGACT ACGATAGCGA TTCTAAAGTT	360
20	ACGAAGACAT TTGAAGAAaT TTTTAGAGAA TATTATTATA ATAACCCGCA ATATCAGAAT	420
	CTTTTTtACA CAGGAAGTAA TCGAATCCT ACTTTAAAG AAACGAAAGG TAAAATTGTC	480
25	CTATTCAATA GAATGGGGGG TACGTACATA AAAAGTGGTT ATGGTGCTGA CACGTCAGGT	540
	ATTCAATGGG CAGACAATGC GACATTTGAA ACGAAAATTA ATAATGGTAG CTTAAATTTA	600
	AAAGTACAAG ATGAGTATAA AGATTACTAT GATAAAAAAG TTGAAGCTGT TAAAAATTTA	660
30	TTGGCTAAAG CTAAAACGGA TAGTAACAAA GACAATGTAT ATGTGAATTT CTTGAGTGTA	720
	GCGTCTGGAG GCAGCGCATT TAATAGTACT TATAACTATG CATCACATAT AAATCCTGAA	780
	ATTGCAAAAA CGATTAAAGC AAATGGGAAA GCTAGAACGG GTTGGCTGAT TGTTGACTAT	840
35	GCAGGATATA CGTGGCCTGG ATATGATGaT ATCGTAAGTG AAATTATAGA TAGTAATAAA	900
	TAAGGATTCA ATAATGATAT TAAGACGAGT ATGAAAATAG TTAGATTCTA ATTATTTTCA	960
40	CTACTCGTTT TTATTTTGAA AATAAGTAAT AATTCAACAA TATTATAAAT TGAACAGATT	1020
	GTTTGTGAAA TTTTGTATAA TATTAAAGTG AAAAAGTGTT ATAAATTGAT AAATATATGT	1080
	AATTAACAAA AACAAATCAT TTAAAAAGA AGAGAGTTGT AAGATGATGa AACGATTAAA	1140
45	CAAATTAGTG TTAGGCATTA TTTTCTGTT TTTAGTCATT AGTATCACTG CTGGTTGTGG	1200
	CATAGGTAAA GAAGCGGAAG TTAAGAAAAG CTTTGAAAAA ACATTGAGTA TGTACCCTAT	1260
	TAAAAATCTA GAGGATTTAT ACGATAAGGA AGGCTATCGT GATGATCAGT TTGATAAAAA	1320
50	TGATAAAGGT ACATGGATTA TAAATTCTGA AATGGTTATT CAACCTAATA ATGAAGATAT	1380
	GGTAGCTAAA GGCATGGTTC TATATATGAA TAGAAATACC AAAACAACAA ATGGTTACTA	1440
55	mTATGTTCGAT GTGACTAAGG ACGAGGATGA AGGAAAACCG CACGACAATG AAAAAAGATA	1500

## (2) INFORMATION FOR SEQ ID NO: 362:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1216 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 362:

TACATGTTTC GGATGCTACT TTATTTAGTT TGAAGGGTGC ATTATGGACG TTAGCGCAAG 60  
 AAGTTTATCA AGAATGGTAT TTAGGATCGA AGTTGTATGA AGATGTTGAA AAGAAAATAG 120  
 CACGAACTAC TTTTAAGACA GGTATATTTT ATCAAGAAAT TATTTTGAGA CCAGTAGATG 180  
 AAGTTAAGGT ACTTCTGAAT GATTTAAAAG GTGCTGGTTT CGAATTAGGT ATTGCAACAG 240  
 GTCGTCCTTA TACTGAGACT GTTGTGCCAT TTGAAAATTT AGGATTGTTA CCATATTTTG 300  
 AAGCTGATTT TATTGCAACA GCAAGTGATG TTTTAGAAGC AGAGAATATG TATCCGCAAG 360  
 CACGACCATT AGGAAAGCCG AATCCTTTTA GTTATATCGC AGCTTTATAT GGTAAATAATC 420  
 GCGATAAATA TGAATCTTAT ATCAATAAGC AAGATAACAT TGTAATAAAA GATGACGTAT 480  
 TTATAGTAGG CGATTCTGTA GCTGACTTAT TAAGTGCTCA AAAAATAGGT GCAACGTTTA 540  
 TTGGAACATT AACAGGTTTA AAAGGTAAGG ATGCTGCAGG TGAGTTAGAA GCGCATCATG 600  
 CCGACTATGT TATTAATCAT TTAGGTGAAC TTAGAGGTGT ACTAGATAAT TTGTAATTTG 660  
 ATTGTTGTTT GACAGCATAA CTTGTAGTGA ATGATTGAAC CAAAGGTTTC ATATTGAGTT 720  
 ACAATGAAAT TAATAATGAA AAAATGCCAA GAAGCAATGG AAGTAATCCA ATGTCTTCTT 780  
 GGCATTTTGA ATTTACATAA ATTGTTTATG ACTGTACCGT CAATTCAGTT GTGAAAATTT 840  
 GATTGTATTC ACCAACTTGT TTAAGTTCAT CAATTATATT GTTTGAAACA GGTGATCAA 900  
 CGGATAAAAT CATTAGCGCA TCTCCGCCG CTTCAGTTCT ACCTAAAGTC ATAGATGCAA 960  
 TGTTGATATT GTATTTACCT AACAATGCGC CAGTTTTTCC TACCATACCT GGAGTATCAT 1020  
 TATGATATGA CACAATTTGA TATTGATTG GCTTTAAGTC TACAGAAAAA TTATTAATTC 1080  
 TAACAATTCT TGGACCGAAA CCTGTAAAGA CAGAAGCGCC AACTTTAACG GAATCGCAT 1140  
 CGNTTGATAG TTCTACCTCT AAGTAGTTAC TAAACCTGT CTCTGCTTTA TTATTTTCAA 1200  
 TATTTAATGT CACTTG 1216

## (2) INFORMATION FOR SEQ ID NO: 363:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 867 base pairs

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 363:

	ATCAAAATAT TTAAATAAT ATTGATGGTC ACATTGTAAA TTTAATAGAA AATAAATTTG	60
10	ATCAAATATT ACAAGAACCA TTAAATCCAT TAAATTATGA TACTGTCAGT GGATTAGCTG	120
	GGATAGGGAG ATATTTGCTA AATAGAGTAG ATGAGAATGA ATTTAATGTT AAAGCATTA	180
	AAAGCATATT AGTATACTTT AAAGATATTC AATATTCTAA AAATAGCTGG GTAGTCCCAC	240
15	AAGATAGTCA ATTTTATAGAG TCTGATAAAA ATTATTTTAC TGAAGGTAAT ATCAATCTTG	300
	GCCTTGCACA TGGAGTGCTA GGACCGATGT CTTATTTTGC ATTATGCGTG ATTAAAGGAA	360
	TTACGATTGA AAATCATCAG CACATATTAA AAGACATGTA CAAATTTATC ATGGACGAAA	420
20	AATTTTGTAA CCACGAAAGA TGGTTCAGC GTTACGATTT AATTTCTGAA CGTAATCATT	480
	TCAATTTTAT TCGGAATGGT TGGTGTATG GCAATACGGG TGTAAATGACG ACGTTGTTTT	540
25	TAATCGGCCA AGCATTACAA GATGATGAAA TAATTAAAAT GTCTAAAAAA GTGATGCTAC	600
	AAGTAGTAAA TGATAAAGAT GAAAATTTAA TAAGTCCAAC TATTTGTCAT GGATTGTCAT	660
	CACAAATATT AATGTTAACA ATTATGAATT TGAATTTTGA ATTAAATGAA GTGTCTGATT	720
30	ATATCACTGT ATTAATAAAT AAAGTATTT CTCATTATAA GGAAGATTAT CTGGTGAATT	780
	TTATAGACAT TAATGAAAAT AAGCAAGATG TATTTAAAAG TAGGAAAGTT GGCCTTTTAG	840
	AAGGTGAATT AGGGGTATC TAACATT	867

35

(2) INFORMATION FOR SEQ ID NO: 364:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 10813 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 364:

45

	TACCTTTTCT TTTAAATCAT TTTATATTTT CCCaCTAATA TCCGcTGtTA ATCaATCctG	60
	ACATCctTGT ATCaCTATGA CAATTAATTG TTAAATACAT GAATTTCTAC ATTTTATGAA	120
50	AAAATCCATT TTTATTACAA TTCAACACTT TATATGACAA CTTCAATTACA GTTACTTTTA	180
	TTGTTGATTG CTTACATTGT TTTCTAAAAA AAATTTGTTA TCATAATTAA CGTTGAATAA	240
	AGAAAAAAT TAAGTTGGGA GATAAAAATG GAATATAAAA AGATACTAAT TCGTTTATTA	300

55

	CAATCGCATG CCGCAGTTAA TTATTATAGT AAAAACCAAT GTACATGGTG GGCATTTAAA	420
	CGTCGCGCAC AAGTCGGTAA ACCTGTTTCT AATAGATGGG GCAATGCTAA AAATTGGTAT	480
5	TACAATGCAC GTAAATCAAA ATATGCGACT GGTTCGTACAC CAAGAAAATT TGCTGTCATG	540
	CAATCAACTG CAGGATATTA TGGACATGTC GCAGTTGTTG AACAAGTATA TAAAAACGGT	600
	AGTATTAAAG TTTCAGAATA CAACTTTTAT CGCCCATTA AATACAATAC ACGTGTACTA	660
10	AGCAAAAAGG CAGCACGTAA CTTTAACTAT ATTTACTAAT CAAAAAATT CTATCACGAA	720
	CGCTTCAATT TCCTGTATGC GTGTGATAGA AGTTTTTATT TTATGAAATT ATATTATTAC	780
15	TTCTACAAAT TTCAAATTGC CGTAATTGAA CGTATATTTT TCCTTCAACT ATTATTTTAT	840
	CTTTAGCATA ATCTATATAT AAAATTTTAT GCTATTATTT AAATAATTTCG CTATAACTTA	900
	ACATACGTTT TCGATATAAA CCTTGTTCTA AATCTCAATA ATTTTTTGCT GTTTTCATCG	960
20	TCATTAGTTA AAAAAATAAT TTAAGTGTG TTTTGAGACC TGTTTAAATT GATGACTCAA	1020
	AACCTTTATC CCCTTTTCAC TCGGTTTAAAT TGCTTTAATA TTTAACACAG TCTCATTGTA	1080
	ATTTTGCTAC TAAGTTTGAA ATATTTTCGAT TCGAATGTGA ACATCATTTA TTATTACTTT	1140
25	CGACAGCTAG AAAAAATATTG TTAAACCAAA ACTATAATTA CCACTTTAT ATAGAACTAT	1200
	ATATAAATTT TAACTTTGAG ACAGAACTGC TAGTCAGGTT TATGAATATA TTTCTTTAGT	1260
	TTACTTGaTA TACTTATTGG TAAATCATT TTTATTGAA GAGCATTTAT AACAAAAAGT	1320
30	ATAATCCCAG TTATAGCGAT TCCATGACCA TTTTATCAG TCAAAGATTC ATCAACAAAA	1380
	ACTTTTTGAT AAATTTATAT ATTTGTATAA TTTATTATGG TAGATATCTA CACCCGTATC	1440
	TATAACAGCT ACAGTTATCA TTGCGTCTTA TCCAAATAGT TTTTAAGAAA TAAATATAAT	1500
35	TCATTAATAT CATGGCTATT TGTAACGTCA GtGCTTaAGA CCTTGTCaAT TATTGTTTTA	1560
	TCTGTTTCTA AGAGCCCCTG ATTTTCTGCA TACATATTTT GTATATCGCT GTTAATCTCA	1620
40	TTTAATTTTT TAAATTTTTT TCTAACCTTT TTCAAAATAA AATGCTCTTT AATTCCATCT	1680
	CTTAAATACA TATAAATCTC TATTATAATC TCCATAAATG TCATATCTCC TCTCTCAAAA	1740
	AACATTATAC TCAATATATT AAAAAATAAGA AAGGTTTTTT CAGAATTTTT AGTATTTTCA	1800
45	GATATTTATT CTTAAGTATC TTATTATTAC TTCTTTTAA CGCTAGTGGa ATAATTAATA	1860
	ATAATTGTAT TTAAATGCAA TGCTGTTGTT CTTTCATACT TACAAGCAAG TCATAAGAAA	1920
	TGAGAAATTA AATTCTTATA TGATGAATAA ATGATGACTG CAGAAGAAAT TGCTAAAAAA	1980
50	GGAGTGGAGT GAGGCATTCT ACTGTTTATA TTGTTAACAA ATATCAAAAT GATAGGAACT	2040
	TGAAGAATCT AAGACTGTTA AAATCGAGAC AAGAAGAAAC TCGTTCCAAA TTTGAATCCT	2100

55



	AAAACCTTCTA TCACGAACGC TTCAATTTCC TGTATGCGTG TGATAGAAGT TTTTATTTTT	2220
	GTTTCATATTA ATTTATCTAA GCGCTACGAT GGAaCTGACTT AAACCTTTTCT TTAAAATCGC	2280
5	TATTCGCCAT TTACTATTGT TGTCTAATTT CTTGTAAAAT ATGTTCCGCT GCTTGTGTAT	2340
	TTGCACGGGG TTCTTTTTTC AAAGCTTCAG CTACTTTAGC AATTTTCATCA CCTTTTGCCC	2400
10	CTACAACGAT AGCTAATGAT TTATATTGTA AACTCATATG ACCTTGTTGA ATACCTTCTG	2460
	ACACAAGCGC GCGACATGCT GCAAAGTTTT GCGCTAAACC AACGGCAGCA ACTACATGAC	2520
	CTAATTCTTG TGCTGACTCT ACATTTAGTA GCTCTAATGA AGCTTTAGCA ATTGGTAATA	2580
15	CTTTTGTACC ACCGCCAACG ATTGCCAATG TCATAGGCAC TTCAATTGTA CCAATCAATC	2640
	GTTGACGATC TTGATCGTAA CGCCATGTAG CAATACCACG ATACTGTCCG TCACGACTCG	2700
	CGTATGCATG CGCACTTGCT TCTGCACCAC GCGTATCATT TCCTGTTGCT AAAACAACAG	2760
20	CATGTATGCC ATTCTAACA CCTTTATTAT GTGTTGCTGC ACGATGAATA TCTACTTGGG	2820
	CCAATACAGA AGCACGTTC ATTCTTTGG CAACCTCTTC TCCAGTTCTC TCGCCCCTTG	2880
	CTAAATCTTT AACATCAATT TCGCCTTGAA CTTTAAACAAC GGACGCTGTT GCATGATTGG	2940
25	ATAAAATACT CATTAAAATG TCGCTTTGCG GAAATTCATT TTTTAAAAAT GCAGTTATGG	3000
	CCTCTAAAAT CGTATTAAAGC ATATTAGCGC CCATAGCATC TTTCGTATCA ACAAATACTT	3060
30	TTAAAGATAG TAACTGTTGC TCAGGAAATG TATCAATCGC TATACGTTGG TAACCACCAC	3120
	CACGCGCTTT AATAGAAGGA TATGCCTCAT CCGCAATTTT ATGAATTTGC TTTTCTAAAG	3180
	CTTTAATGTC TGCTGATAAT TTTTCAGTAT CGTCAACGCC ATCAAAGACG ATTTGACCTA	3240
35	TCATAATACG TTCAGAAGAT ACCGTTTTAA ATCCGCCAGT CTGATTCACT AGCTTTGCAC	3300
	CATAACTAGC TGCAGCGACA ACTGAAGGCT CTTCCACCAT CATAGGTACA ACATATGCCT	3360
	TATCGTCCAC AATGATATTC GGTAAATAATC CAACGGGTAA TGCACCTTGC GCGATGACAT	3420
40	TTTCAATTAA ACTATTGGCT ACTTCTTCAT CGATTAAATGG ATGATTCAGT AAAATGTCGA	3480
	ATTGTCTTTC TGATAACCAT TGCTTATCAA CCAATTGTTG TAACTTTTCT TTACGAGATA	3540
	AATGTCGAAA ATTCTTATCT AAATTTTGCA TGGACGTACT CCTTTTACTT CACATAATTT	3600
45	TTAACATTTT AATCACTACT ATTTTACCA CAAATAACG TCATTTCGTCT TAAAATTCAA	3660
	TTGAATAATT GTCGTTTTGA CTTTAAAATA AAACAAGGTA AATTAAAACG CTTACAAGAA	3720
	ACGACAAATc ATTTTAAAT TTAGTATATT TCTTTGTATA AAATTAGCAT ATTCTGATAT	3780
50	GATACAAGTG TTGCTTTTAT AAATTTGAAA GGATGTAAAA CCTTATGACA ATAGGTATCG	3840
	ATAAAATAAA CTTTTACGTT CCAAAGTACT ATGTAGACAT GGCTAAATTA GCAGAAGCAC	3900
55		

	CTGTAAACCA AGACATCGTT TCAATGGGCG CTAACGCTGC TAAGGACATT ATAACAGACG	4020
	AAGACAAAAA GAAAATTGGT ATGGTAATTG TGGCAACTGA ATCAGCAGTT GATGCTGCTA	4080
5	AAGCAGCCGC TGTTCAAATT CACAATTAT TAGGTATTCA ACCTTTTGCA CGCTGCTTTG	4140
	AAATGAAAGA AGCTTGTTAT GCTGCAACAC CAGCAATTCA ATTAGCTAAA GATTATTTAG	4200
10	CAACTAGACC GAATGAAAAA GTATTAGTTA TTGCTACAGA TACAGCACGT TATGGATTGA	4260
	ATTCAGGCGG CGAGCCAACA CAAGGTGcTG GCGCAGTTGC GATGGTTATT GCACATAATC	4320
	CAAGCATTTT GGCATTAAAT GAAGATGCTG TTGCTTACAC TGAAGACGTT TATGATTTCT	4380
15	GGCGTCCAAC TGGACATAAA TATCCATTAG TTGATGGTGC ATTATCTAAA GATGCTTATA	4440
	TCCGCTCATT CCAACAAAGC TGGAATGAAT ACGCAAAACG TCAAGGTAAG TCGCTAGCTG	4500
	ACTTCGCATC TCTATGCTTC CATGTTCCAT TTACAAAAAT GGGTAAAAAG GCATTAGAGT	4560
20	CAATCATTGA TAACGCTGAT GAAACAACCTC AAGAGCGTTT ACGTTCAGGA TATGAAGATG	4620
	CTGTAGATTA TAACCGTTAT GTCGGTAATA TTTATACTGG ATCATTATAT TTAAGCCTAA	4680
	TATCATTACT TGAAAATCGA GATTTACAAG CTGGTGAAAC AATCGGTTTA TTCAGTTATG	4740
25	GCTCAGGTTC AGTTGGTGAA TTTTATAGTG CGACATTAGT TGAAGGCTAC AAAGATCATT	4800
	TAGATCAAGC TGCACATAAA GCATTATTAA ATAACCGTAC TGAAGTATCT GTTGATGCAT	4860
30	ATGAAACATT CTTCAAACGT TTTGATGACG TTGAATTGGA CGAAGAACAA GATGCTGTTC	4920
	ATGAAGATCG TCATATTTTC TACTTATCAA ATATTGAAAA TAACGTTTCG GAATATCACA	4980
	GACCAGAGTA GTCGGTGTAT TTAAACACA TATAATAAAA CCTAAAAGCA GCAGTAAGAC	5040
35	CACCTTCTAAT TGAAATCGTC TTAGTCTGT TCTCTATTTA TAACACTTCG TATTGAATGA	5100
	ATTCATTATG CCTATTTGAC ACATTATTGA AGTTTTCCCTA ATGCCTGGAT CCTTTATACG	5160
	TTACGGCTTC GTGCTATGTT TTGGTACATA AAGCTTTGAC ATATCGATAT TCTCCAACCTC	5220
40	TAACAGCTTA ATTTTATTAT TAATCGTTCC ACCGAACCCT GTTAAGCTAC CCGTTTACC	5280
	GACAACACGA TGACATGGCA CGATAATAGA TAATGGATTA CTTCCGACTG CACCTCCAAC	5340
	CGCTTGGGCT GACATTTTTC GCTTGTTAAG CAGCTTGCCCT ACTTTTTTGG CAATAGCACC	5400
45	ATACGTTGTT AGAGTCCCAT AAGGAACCTG TCTTAATTCA TTCAAACAC ACTGTTGAAA	5460
	ATGACTACCT GTTGGCTTTA AAGGTATTGT GATTTTCAGGA TTGTCACCTT TAAAATACGC	5520
	GTCTAACCAC TGTGTCGCCT CTCTAAATAT CGCTAAAGAC GTATTTTCTT CCCTAGTACC	5580
50	ATCACCTTGT TGATTTTCAA ACAAACAGC GGTCAACTT ACCCCATCAC TCAAAAGTTC	5640
	CAATCGTCCT ACAGGCGAAT CATAGTAACT CTTATACTCC ATAAAAATTC CCCCTTTTTC	5700

55

	ATAAGTCGTC AATTACGTAT ATAAACACGT AATACCAGCT ATCACTTTGC TGCAATATAC	5820
	AGTTACATAT CTTACTACAC GTGCTAACCT CTTACTTTGT AAACCAAATC TTAAATTAAA	5880
5	ATATTGAAAA TGCAATGAAT CCTTAATATT TTATTAAACC TATAATTACT TATTAAAAAT	5940
	AACACACAAT ATTCATAAAG TTTTAAAAAT ATTCTGTTTT ATCACCTACT ATTAGTGGAA	6000
10	AAGTACAATT GCAATTGTAT ATAGTTTGCA TAACGCTTCA AAAGTAATTT CTTTTTTGTT	6060
	TAGTTCAAAA AAATTTAGAG GTGATGTTAT ATGAATAACG GTTTTTTCAA TAGCGACTTT	6120
	GATTCAATTT TTCGAAGAAT GATGAAAGAT ATGCAAGGTT CAAATCAAGT CGGAAACAAA	6180
15	AAGTACTATA TTAATGGTAA AGAAGTTTCA CCTGAAGAAC TAGCGCAACT CACACAACAA	6240
	GGTGGCAATC ACTCTGCTGA ACAAAGTGCG CAAGcTTTTC AACAAGCAGC ACAAAGACAA	6300
	CAAGGGCAAC AAGGTGGCAA CGGCAATTAT TTAGAACAAA TTGGTCGTAA CCTTACGCAA	6360
20	GAAGCACGTG ACGGTTTATT AGATCCAGTC ATTGGTCGTG ATAAAGAAAT TCAAGAACT	6420
	GCTGAAGTTT TAAGTAGACG AACTAAAAAC AATCCTATAT TAGTTGGAGA AGCTGGTGT	6480
	GGTAAACTG CGATTGTTGA AGGTTTAGCA CAGGCAATCG TTGAAGGAAA TGTACCAGCA	6540
25	GCAATCAAAG ACAAAGAAAT TATTTCTGTA GACATTTTCAT CATTAGAAGC TGGAACGCAA	6600
	TATCGTGGTG CTTTGAAGA AAATATTCAA AAATTAATCG AAGGTGTTAA ATCTTCACAA	6660
	AATGCCGTAC TATTCTTTGA TGAAATCCAT CAAATTATCG GTTCAGGTGC CACAGGAAGT	6720
30	GATTCAAGTA GCAAAGGGTT ATCTGATATT TTGAAACCTG CATTAAAGTCG TGGTGAGATT	6780
	TCTATTATTG GTGCAACAAC ACAAGATGAA TATCGAAACA ATATTCTTAA AGATGCTGCA	6840
35	TTAACGCGCA GATTTAATGA AGTGCTTGTT AATGAACCAA GCGCTAAAGA TACTGTTGAA	6900
	ATTTTAAAAG GTATTGCGGA AAAATTCGAA GAACACCATC AAGTAAATT ACCAGATGAC	6960
	GTATTAAAAG CATGTGTTGA CTTATCAATT CAATATATTC CACAACGATT ATTACCAGAT	7020
40	AAAGCAATCG ATGTGTTAGA TATTACAGCA GCACATTTAT CTGCGCAAAG TCCAGCTGTC	7080
	GATAAAGTTG AACTGAAAA ACGAATTTCT GAATTAGAAA ATGATAAACG TAAAGCAGTA	7140
	AGTGCTGAAG AATATAAAAA AGCTGACGAC ATTCAAAATG AAATCAAATC ATTACAAGAT	7200
45	AAATTAGAAA ATAGTAATGG TGAACATACT GCTGTTGCTA CAGTTCATGA TATTTAGAT	7260
	ACTATTCAAC GATTAAGTGG TATTCCAGTT TCTCAAATGG ATGATAACGA TATTGAACGT	7320
	TTAAAAAATA TTTCTAATCG TTTAAGAAGT AAAATCATAG GTCAAGATCA AGCTGTAGAA	7380
50	ATGGTTTCAC GTGCAATTCG CCGTAATCGT GCTGGGTTTG ATGACGGCAA CCGTCCAATT	7440
	GGCAGTTTCC TATTTGTTGG CCCTACTGGT GTTGGTAAAA CAGAGCTTGC TAAACAATTA	7500

55

	GACACAACAG CTGTTTCAAA AATGATTGGT ACAACTGCTG GTTATGTTGG TTATGATGAC	7620
	AATTCAAATA CGTTAAGTGA AAAAGTACGC CGTAATCCAT ACTCAGTCAT TCTATTTGAT	7680
5	GAAATCGAAA AAGCAAATCC ACAAATTTTA ACATTGTTAT TACAAGTAAT GGATGATGGT	7740
	AATTTGACTG ATGGTCAAGG TAATGTCATC AACTTTAAAA ATACAATTAT TATTTGTACA	7800
	TCAAATGCTG GCTTTGGCAA TGGCAATGAC GCTGAAGAAA AAGATATTAT GCACGAAATG	7860
10	AAAAAATTCT TCCGCCCTGA ATTCTTAAC CGCTTCAACG GCATCGTTGA ATTCTTACAT	7920
	TTAGATAAAG ATGCATTGCA AGATATCGTC AACTTATTAT TAGACGATGT ACAAGTTACA	7980
15	TTAGACAAAA AAGGTATTAC GATGGACGTT TCTCAAGATG CGAAAGATTG GTTAATTGAA	8040
	GAAGGCTATG ATGAAGAATT AGGTGCACGT CCATTAAGAC GTATTGTTGA ACAGCAAGTA	8100
	CGTGACAAAA TTACAGATTA CTATTTAGAT CATAACGACG TTAAACATGT GGATATAGAT	8160
20	GTTGAGGATA ACGAATTAGT CGTAAAAGGT AAATAACGAC ACTTTAACAT ATCGCGCATC	8220
	AAAAATGAGC ATCAGGTGCG CCTTGCCTGT GCTCATTTTT TTAATTATTT CCCTGGAAAA	8280
	TGATTCGCTG TGTGCTGTTT TGTTCACAA CAATCACGAT TAATGTCACA TGTACCACAT	8340
25	TTTCCTTGTT TTGAACGCTT GAAAAATTTT ACTAGTGTAT ATAAGGCATA TCCGAAAATT	8400
	GCTAAAAAAA TTAAATGTT AATAATGACT GACACTTTAA CCACTCCTTA AACAAATAAA	8460
	TGTCCGACTT GATAAAAAAT GAATGTTAAG ACATATGCAG TGAAGTAGAG ATAGGCAACT	8520
30	GCAAGTGCCG TCCATTCCA TGAATAAGTC TCTTTACGGA TTGCTGCTAC TGTAGAAACA	8580
	CAAGGAATAT ACAATAGTAT AAATATCATA AATGCATACC CAGATAGCGG TGTGAATTGA	8640
	TTTTGAATCA CATTAAACAAG GCCTGCATCA CCTGATGAAT AGATAATCGC CATCGAACTT	8700
35	ACGATAACTT CTTTGTCTAA AAATCCTGGC ACTAACGTAG CACCTGCTTG CCATGTTCCA	8760
	AATCCGAGCG GTTGCACTAA CATACCAAAG AACTACCAA CCATATGTAA AAACTTTGA	8820
40	TTGATATTCA CATTGATACC ATGTGGTCCT ACATACTTA ATAGCCAAAT GACTACTGAG	8880
	CCGCCAAAAA TAAATGTACC TGCTTTACGA ACAAAGCCCT TAGCCTTTTC CCAAGTACTA	8940
	CGCCACAACG TTTTAATGGA AGGCACACGG TATGTTGGCA ATTCCACAAT AAAGATTGCA	9000
45	TTATCATTTT TTAATAATCGT CTTAGTAAGT ACTGTACTGA CTAAAAATGC CATAATAATA	9060
	CCTAAAACAT ACAGGCTTAA TACTACTAAA GATTGATTCT CTTTGAAAAA GATACCTACG	9120
	AACAACGCAT AACTGGCAG TCTAGCAGAG CATGACATGA ATGGTGCAAT TAATATCGTT	9180
50	GTAAACGCT CTTTTTCATT TTCAATACTG CGCGCAGCCA TAATACTCGG TACATTACAA	9240
	CCAAATCCGA TAATCATTGG TATAAAAGAC TTCCCGCTTA AACCGAACGA TTCCATAATA	9300
55		

	AAAAAGAGCA CAACAATTTG TGGTACAAAG ACTAATACTG ATCCTACACC AGCAATAATG	9420
	CCATCTGTAA TTAAATCTTG TAAAAATGGT ATAACACCAA GATAATTCAT AATCGTCTTC	9480
5	ACACTATCTG TAAATGTACC ACCTATAAAT GCATCGAGTT GATCCGACAA AGGTGTGCCA	9540
	ATCCATGTAA ATGTAGTTTG AAAGATCAAC CACATAATTG CTAGAAAGAT AGGCATCCCT	9600
	ATATATTTAT GTGTTAATAT CTTGTCTATT CTAGAGCTGA AATATTGCTT ATCTTCATCT	9660
10	GGATACGTTA CCACGTCTTG CAATAACGTC TCAATATAAT GATTGCGTAT ACGCTCCATC	9720
	TCTCGACGAA CAGATACAGC CCCTACTTGT TCAGCAACTT GATCACGTAA ACTCGACAAT	9780
	TTATTTACAA CCTCTGAATT AAGTTCGTTT GCAATTTGCA TGTATTTTAA TAAGAATTGA	9840
15	ATCGCAATAA ACCTAGCTTG ATACTTATCA TGAGATGTCT CTGTCATTAT TATTTGACAC	9900
	ATATTTTAA TTGTCTCTTC AATCTTCTCA CCATAATTGA TTTTAAAATG CGGTTGATAC	9960
20	CCTTCCCCTA GATGCTTTAT TTCGCCAAGT AAATATTTTG TTCCTTTGCC TGTACGTGCC	10020
	ACAACGGAA AAATAGGTGT TTTTAACTTT TTCATCAATT TATGATAATC GATTTTATC	10080
	CCGCGCTTTG TAGCTACATC AATCATATTT AATCCGATGT ATATTGGTTG ATTAAGTTCT	10140
25	AACAATTGTA CTGTTAATTG CATATTTCTT TTTAGTTGAC TCGCATCAAC AATGTTAATG	10200
	ATTCCTGAAA ATGAATCGTT TAATAAATAG TCTGTCACTA CAGTTTCATC TTTAGAAATC	10260
	GGCGATAAAT CATATGTACC TGGTAAATCA ATTAATTGTC CTACATTTTC TTTAAGTTTC	10320
30	CCTACTTTTT TCTCTACCGT TACGCCACTC CAGTTGCCTA TATATTCATA CGAACCAGTT	10380
	AAAGCGTTAA ACAAAGATGT TTTACCAACA TTAGGATTTT CTAAATACA ATAATTTTCC	10440
	ATTCGTCCGG CTCCTATTCT TCTAATGCAA TAGAACAAGC ATCGCAATGT CTAATACTTA	10500
35	ACTGTTGTCC GTTTACTTCA ATAATACATG GCCCTTTAAA TAAACATTTT TGTTTAATCG	10560
	TTATGATAGC GTCATCTGTT AACCCAAAGG CACTTAGACG ATACAACATA TTCTCATTAG	10620
40	CAATATCCAT TCGCTTTATT TTATAAGCCT TATTCATTTT ACCATTTTTA ATGTTTAACA	10680
	TACTATTTTG CTCTCCTATT AGAAATAATA ATCATTATCA CTTAAAAATC ATAACCCTTA	10740
	AAATTGTAGC TCGCAATACT TTATTTAAAT AATTTTCATT TTTCATGTAA AATTTGTGAC	10800
45	ATTGCAAAAA TGT	10813

(2) INFORMATION FOR SEQ ID NO: 365:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6804 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 365:

	TGATGGATTA GCAGACATTT TACGAGCGAA TGGTTTCAA GTGTTTGGTC CAAATAAGCA	60
5	AGCAGCTCAA ATCGAAGGCT CAAAATTATT TGCTAAAAAG ATAATGGAAA AATATAATAT	120
	TCCAAC TGCT GATTATAAAG AAGTTGAGCG AAAAAAGGAT GCTTTAACAT ATATTGAAAA	180
	CTGTGAATTG CCCGTTGTTG TCAAGAAAGA TGGGTTAGCT GCTGGGAAAG GCGTTATTAT	240
10	TGCAGATACT ATTGAAGCAG CCAGAAGTGC TATTGAGATT ATGTATGGTG ATGAAGAAGA	300
	AGGTACTGTT GTATTTGAAA CGTTTTTAGA AGGTGAAGAG TTCTCGCTAA TGACATTTGT	360
15	TAATGGTGAT TTAGCAGTAC CTTTCGACTG TATTGCACAA GATCATAAAC GCGCATTTGA	420
	TCATGATGAA GGACCAAATA CTGGTGTTAT GGGGGCTTAT TGTCCmgTAC CACATATTAG	480
	TGACGATGTT TTAAmACTTA CAAATGAAAC AATTGCACAw CCCATTGCAA AGGCAATGCT	540
20	TAATGAAGGT TATCAATTCT TCGGTGTATT ATACATTGGT GCTATTTTAA CTAAAGATGG	600
	TCCAAAAGTA ATAGAATTTA ATGCCCGTTT TGGTGATCCT GAAGCTCAAG TATTATTAAG	660
	TCGCATGGAA AGTGATTTAA TGCAGCATAT TATTGATTTA GATGAAGGAA AACGTACTGA	720
25	ATTCAAATGG AAAAAATGAAT CTATTGTAGG GGTGATGTTG GCATCAAAAG GATATCCTGA	780
	TGCATATGAA AAAGGGCATA AAGTAAGTGG CTTTGATTTA AATGAAAAC ATTTTGTTAG	840
	TGGATTAAAG AAGCAAGGTG ATACCTTTGT TACTTcAGGT GGTAGAGTTA TACTTGCCAT	900
30	CGGAAAAGGT GACAATGTAC AAGATGCACA GCGAGACGCA TACAAAAAG TATCACAAAT	960
	ACAAAGTGAC CATTTATTCT ATCGTCATGA CATTCGCAAT AAAGCACTAC AACTTAAATA	1020
	AGTAAATTTA AAATACTAAG aTTAGCTATG AACGAATCTA TAACGATAGA TTTTTTCATA	1080
35	GCTTTTTTAG TTGTAGAGTC TAGGACATTG ATTCTGTAC CAAATTTGTG ATTATGCATA	1140
	TGTAATACAA AAGAGGCGCC ACAACATGTT TGGATGAACA AAATAACATG TTTGTGGCAC	1200
40	CTCTTTTGTT TAGTATGGAA TAAATGGTTT TCTTTTCTA TACAATGAAT TTCTAATTTA	1260
	GTATCTATAC AATTATGGAT AAAATTTAAC CTACACGACC AAGACGAACA TCATCTATGC	1320
	CCGTGATGGG TAAGGTGATT GAACAATAAT ATGCCATAGT AATAATGGCA ATTAAACTA	1380
45	TAATAAAGAT TATATCTTTA TATGAGAAAG GTACGTTGTA ATAGTAAGTA CGAGGACCAT	1440
	CTCTAAATCC TTTCGACTCC ATCGCAACTG ATAATTGATG TGCCTTTCTA ATATTTTGGC	1500
	TTAATAGAGG TATAATTAAA TGCTTAAATC GCTTTAACCC TCTATAATTT GCCGCGTCTA	1560
50	TCATCTGATA GCGCATTTTT AAAGATCTGC GAAGcTGTA TAAAGAACTA ATCATTAAG	1620
	GTATCATACG AATGGCAGCC ATGAATGCAT AAGCAACTTT TGATTTAACC TTTAAATGTT	1680

55

	ATGAAATAGC AATGGTTCTT AATGATACAT GTAAACCACG AACTAAACTT TCTGTTGTAA	1800
	TATGGATAAA TCCGAATTC AAAATTGTAT GGCTACCATT CCCGTATAAA ATCATGAACA	1860
5	GGGAAGAGAG TAATGCAAAG CCAATACTTA TAGTTATAAA AATTGCTGTA ATTTTAACT	1920
	GAGTACCATT AAACATCAAT AAGAAACTA ACATTAAGAT AGTGATATAA AGCATAAAAT	1980
	CGAAATTATG CACAAATATA ATAAAGAAAA ATAGTATAAT TCCAAGAAAT AGTTTCGTTA	2040
10	TAATGTTGAC ATCATCAACA AATGATTGCC GAACTTTCCA TTGCTCATAC ATTCGTATCA	2100
	CCATCACAAT CTAGTAACGC ACCATCTGAA ATTTTAAGTC TTCTTGATGG ATAACGTTCA	2160
	ATTATTTTCAT CGTCATGTGT AaCCATGACA ATACTTTGTC CCAAATTAAT TCGCTTTTGG	2220
15	AAAAGTTTGA TCAACTGGAA TGTATTATGG CTATCAAGTC CAAATGTCGG TTCATCTAAA	2280
	AAGATAATAT CAGCTTTAGA ACTTAGTGCG GTAgcTACGC TAAGGCGTCG TTTTGTACCA	2340
20	ATAGACAACT CATAAGGATG TTGATCTTTT ACATTTTGTA AATCTAAAAG TTTTAAAAGT	2400
	TGTATCGTTT CATCATCACT TTGATCTTTA GAAAGGTGAT TAAATGAAT GTTAATTTCA	2460
	TCATAAACCG AATTTGTTAT AAATTGTAAT TCTGGGTTTT GATAAACTAG GTACATGTGT	2520
25	TTTGCTGCAT GTTTAATTTT TGTAAACGC TGATTTTCAA AATAAACATC ACCTTGATAT	2580
	TTAATCAATT GCATAATTGA TTCAAGCAAG GTTGTTTTAC CACTACCATT TGCCCCTGTA	2640
	ATTGTAATCC ACTCACCTAG ACCAATTTCT AAATCTGAGA ATGAGAGCAA TGTTGATTTA	2700
30	CCGCGAATAA TACGTCCATT TTTAAATTGT AATAAGTGTG AGTTTGTTGT TGGAAAGTCA	2760
	ACACGACTTG GTGCGAATTC CCATGCACGT GGATGCCACA CACCATATTC ACTGAGTAAA	2820
	TGAACATACT TCTGTAATAT GATTTCAAGG CATTTCATCGG CAATGATAAT TCCGTTATAA	2880
35	TCCATCAAAA TGACGCGGTC GACATGATTC CAGATGTGTT TAACTTTATG TTCAACGATT	2940
	ACAACCGTTT GATCTTCCCA AAGTTCAATT AGTTTAGTCC ATAAATCTTC TGTTGCTTGA	3000
40	ACATCTAACA TTGCTGTCGG TTCATCTAAA AACAATGTTT TTGATTGTTG AAGAATGGTT	3060
	TCAACAATTG CCAATTTCTG TTTTCATCCCG CCACTTAAAT CTTTGATATA CGTTTCAGGG	3120
	GTAACATTTA AATTGACCAT ATTTAAAGCA TTGATAATTA ACGCATCCAT GTCTTCACGT	3180
45	GGTAATTGTC TATTTTCTAA AACGAATGCA AGTTCTTCGT ATACTTTTGG CATAAAAAC	3240
	TGGCTATCAG GGTCTTGGA AATAACGCCA CTTAATGGGT CAACGATTAG TTCATCATAT	3300
	TTCATAGGTA ATTCAATTAA ATTAGGAACA ATACCACTTA ATACATTCAG AAGTGTAATT	3360
50	TTACCGCAAC CAGAAGGACC GAGTAAAAGT ACTTTTCTT TGTCTTGAAT AGTGATATTT	3420
	AAATGATCGA AAATTTTACG TTGACCACTT GGATATTTTA ATCGTAAATC ACTTACTTTT	3480

55

	ATTTTGTTAC GCCTGTCTTA TCTAAAGCTT TTAATAAAAG GTAAGATAGG ACGCCGGCGA	3600
	CTACTGCACC ACTAATTAAT CTAAATACGA TGAATAATGT TAAGTTCCAA CCTGCAACTT	3660
5	CATTTAAATA ACCATAGAAA TAATCTATCG GGAAAGCCGC GATTGCTGTA CAAAAACCTG	3720
	CTAACATAGC TACCATAACT GAACGTGATT GATATTTAAA AATTGCAAAG ACAAGTTCAC	3780
	ACGCTAAACC TTGTATAAAA GCGTAAACGA TTGTCGGAAT ATCGAAACGA CCCATAATGA	3840
10	TAGTTTCGCC GGCACCTGCA GCAAATTCAG CCAGTAAAGC AATACCTGGT TTTGGAATAA	3900
	TTAGATAGCA GACAATCGCT GCCATGAACC AAACCCCGTT TGTTAATGT TCGAGGTGAA	3960
15	GGCCTGTAGC TTGCACACCA TTGTAAACAA ACCACCATAA ATTGTAAATA ACTGCGAATA	4020
	CTACTGAAAT AAGTACGGT ACTAGTATTT CAGATAGCTT TAAACCTTTT GACATTTTTA	4080
	CATCCTCCTA ATAAAAAAC GCACAACCAT CCATAGGAAA GTTATGCGTT CACAATATAT	4140
20	ATTAGTAAAA CATATGTATA GTAACACTTT CCTACGCTAG TTCAAGCTAG ATCAGGTTCA	4200
	AAGGGTTTGA GGGCAAGCCT CATCTCAGTA TAAACACCC CTAGTGTGTG CGATTTATTT	4260
	AATTAATTAT ACTGTAAGAC GTTTGTAAAC TTATGTCAAT AGGTTGTCTT CATGAAATTT	4320
25	CGTTTAATTC GATTTAAAT TTATAATATT AGCATTGGAT TTAAATTGAA GATGTAGTAG	4380
	GAATGTTAGT AATTAAAGAT ATAAAAATAT GTGACATGTA ATAATATTGA GCTGATAAAT	4440
	GAAGAGGGAT ACTTATCAAT CATACTCTT TAACAACAGT GAAGAACCCG TGCATAATGG	4500
30	CTTACGAATT ATAGTTTATA AGGAAGAAGA GGGATACATG CGCCGAGCAC ATGCATAAAA	4560
	GCCCCTAACA ACTAAAAGTT GTAAGGAAGG AGAGGGATAC ATGCGCCGAG CACATGCATA	4620
35	AAATCCCCTA ACAACTAAAA GTTGTAAGGG GATTTAAATT AATTTAGTGT ATCTTGGATA	4680
	TCTTGTTTTG KTTGaTTAAT ATCTCTGTT TTTCTTCTT TTTTATCTTT TAATTTTCT	4740
	TCAACTTCTT TAGCTTTTTC TGCTGCTTTT TTATTTTGAT TTTCATTAGA CATGATTAAT	4800
40	TCCTCCCAA TTGGATAATT ATTTATATAT AAATCTTACC CGGTTGTACT TTCGTTAAAC	4860
	TTTTCTAAGT CTATAGCACT ATTTATTCAT TTATCTAAAG ACAACAACAT TAGATTAATA	4920
	TATAATGATT TTGAGGTGAA CATAATGTCT TTTCTTAGGA AACACGCCGA AATTATTTTT	4980
45	AGCTATTTAA TCGGTmWCGT TcACTCTTCA CTGGkcTCAT TATTTTAATT AACTTGCCAT	5040
	TAATTAAACA ATTAAATGGT GGTAAAAAAG TTGATACACA TGTCATAAT GTGTGGGAAT	5100
	TTCTGAATGC ATTTTTCAGT GAAATTATTA AAGTAATGAG TCGATTTATA GGTAATTTCC	5160
50	CcATAGTTAG TGCAATTGTG ATAATTATAT TCGGTATTTT AGTTATGTTG ATTGGTCATA	5220
	CATTACTTAG AACTATTAAG TATGACTATG ATATTCTAT CTTTTTCTTA GTTATCGGTA	5280

55



TTTTCATTAT TCCATTTACA ATTCATATAG GATATATCGT CTATAAAGAT GAATTGAATC 5400  
 AGGAAAATGT AAAAAATCAT TTCATGTGGA TAATTGTGAG TTATGGTATA AGTTACTTAA 5460  
 5 TTACACAAAT TGCATTGTAT GGCAGAATTG ATGCTAATGA AATAGAGTCA ATTGATATCT 5520  
 TAAGTGTCAA TGCTTTCTTT ATAATTATGT GGTACTTGG TCAAATGGCT ATTTGGAATT 5580  
 TCTTGTCTTT GCGCCGAGCT TTACCTTTAA CAAAGCAAGA ATTAGGTGAA GAGGAGCCAG 5640  
 10 AATTATCAAG AACAAGTAAA GGAATGTCA CGAATCAAAC TAAATTCAC TTGAAACAAC 5700  
 TCCAAGATAA GACTACAGAA TATGCACGTA AGACAAGAAG AAGTGTCGAT TTAGATAAAA 5760  
 TTAGAGCTAA AAGAGATAAA TTCAAAAAGA AAGTTAATGA TATTATCGAT ATTCAAGAAG 5820  
 15 ACGATATTCC TGATTGGATG AGAAAACCGA AATGGGTAA ACCAATGTAT GTCGAACTAT 5880  
 TTTGTGGTGT CGTCATCTTT TTATTCACAT TTTTAGAATT TAATAATCGT AATGCATTAT 5940  
 20 TTGTATCTGG TGATTGAAA TTATCACAGA CACAATATGT TATTGAATGG GTTACATTAT 6000  
 TAATTCGTGT ATTCAATTAT ATCGCATATA TCGCTACAAC GTTAACTTTC CACTTGAAAG 6060  
 GTAAGTTTAA TTATTACAA TTATTTATGG GGAGCATTTT ATTCTTTAAA TTGTTAACGG 6120  
 25 AATTTATAAA TATAATGATT CATGGACTAT TACTTTCAGT GTTCATTACG CCAACATTAC 6180  
 TATTAATGTT ATTGGCAATC ATCATTTCTT ATTCGTTACA ATTACGAGAG CGACCATAAT 6240  
 TAAAAGCATT ATAAAAGTAC TATCTATTAA ACATTTTGAT GTGTACGCTA TAAGTTAGAT 6300  
 30 ATATCTCTAA CTTACTTAGA TACAGGTCAA TGAAGTTTAT GGATAGTACT TTTTTTGTA 6360  
 CTAGATTTGA TTGATTCAGG TGATGTGAAT TAAGTATTGA TAATTGTATA CAAAGTTTAA 6420  
 GTGCAAATAA AATAGTTGAA AAGTTATCCA TTTGTAAAAT CAAGAAAAC AGTAAATAGT 6480  
 35 TGAAGCGACT TATGGAATTT GCGAAACGAT ATATAGTATT TCCTTTGTAG AAATTTTACA 6540  
 TATATCATT CAAATTACTAA TTTGTTAAAA TCAACAGTAA GATTAGAAGT AGATGATATT 6600  
 GAAATTTGGC AAACAATTATA ATCTATATAA AACTACAACG AAACACAGAA AGGAAGTTGT 6660  
 40 CAGATGAAAA TAGCAACTCT GAACAAAGGC AAAGAAACAA AATATTTTAA TGGATATCCT 6720  
 TTAATTGAAG AAGAGGATAT CTATTCACAA GATCATTTAA AAGAAGGAGA TATTTTTCAA 6780  
 45 ATTGTGACTG ATAAATCACA ATAT 6804

(2) INFORMATION FOR SEQ ID NO: 366:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1717 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 366:

	aaaAGAGACG CATTaAAACA AGCTATACAA ATTATCGATA AATTAACATG GGGTGTtTAG	60
5	TGGTGGTTAA AGAAATTTTG AGACTATTAT TCTTACTAGC GATGTATGAG CTAGGTAAGT	120
	ATGTAActGA GCAAGTATAT ATTATGATGA CGGCTAATGA TGATGTAGAG GCGCCAAGTG	180
	ACTTTGAAAA AATCAGAGCT GAAGTTTCAT GGTAAATAGCT ATTATCATTt TTGAATTAAAT	240
10	TATATTAAATG TGTGTAGCAA TAGCACTGGA GGTGTTGTAA ATATGTGGAT TGTCAATTCA	300
	ATTGTTTTAT CTATATnTTT ATTGATCTTG TTAAGTAGCA TTTCTCATAA GATGAAAACC	360
15	ATAGAAGCAT TGGAGTATAT GAATGCTTAT CTTTTCAAGC AGTTAGTAAA AAATAATGGT	420
	GTTGAAGGTT TAGAAGATTA TGAAAATGAA GTTGAACGAA TTAGAAAAAG ATTCAAAAGC	480
	TAAAGAGAGG CGTTGGCTTC TCTGCTCTAT CyAAAATAAT GAAAGGAGCC saACATGTTA	540
20	GaCmAAGtCA CTCAAATAGA AACAATTAAA TATGATCGTG ATGTCTCATA TTCTTATGCT	600
	GCTAGTCGTT TATCTACACA TTGGACTAAT CACAATATGG CTTGGTCTGA CTTTATGCAG	660
	AAGCTAGCAC AAACAGTTAG AACTAAAGAA GATTTAACTG AGTACAATAA AATGTCTAAG	720
25	TCTGAACAAG CCGATATAAA AGATGTTGGC GGATTTGTCTG GTGGATATTT AAAAGAAGGC	780
	AAACGGCGTG CTGGTCAAGT CATGAATCGT TCAATGCTAA CACTTGATAT CGATTATGCA	840
	GCCCAAGATA TGA CTGACAT ATTATCTATG TTTTATGATT TTGCATATTG TTTATATTCA	900
30	ACACATAAGC ATAGAGAGAT AAGTCCAAGA CTGCGTTTAG TGATTCCTTT AAAACGAAAT	960
	GTAAATGCAG ATGAGTATGA AGCTATTGGG CGTAAAGTCG CAGATATCGT TGGCATGGAT	1020
	TACTTCGATG ATACAACTTA TCAACCACAT AGGTTAATGT ATTGGCCTTC AACTAGTAAC	1080
35	GATGCGGAAT TTTTCTTTAC CTATGAAGAT TTACCTTTGT TAGACCCAGA TAAATATTA	1140
	AATGAATATG TTGATTGGAC TGACACATTA GAATGGCCAA CGTCTTCAAG GGAAGAGAGT	1200
40	AAGACTAAAA GATTAGCAGA TAAGCAAGGC GACCCAGAAG AAAAGCCGGG AATTGTTGGT	1260
	GCATTTTGTA GAGCCTATAC GATAGAAGAA GCTATAGAAA CTTTTATTCC TGATTTATAC	1320
	GAAAAACATT CTACTAACCG TTATACCTAT CATGAAGGTT CAACTGCAGG TGGATTGGTG	1380
45	TTATACGAAA ATAACAAGTT TGCCTATTCT CATCATAATA CGGATCCCGT AAGCGGTATG	1440
	CTTGtGAACA GTTTTGATTT AGTACGCATA CACTTATATG GTGCTCAAGA TGAAGAACT	1500
	AAAACAGATA CTCCGGTTAA TCGACTACCT AGTTATAAAG CAATGCAGCa AAGAGCGCAA	1560
50	AATGATGAGG TTGTTAAAAA GCAATTAATT AATGACAAAA TGTCTGATGC AATGCAGGaT	1620
	TTCGATGAAn GAGAAAATAG CGATGATGCA TGGTCTGAGA CGTtnGAAAT TACTTCGAAA	1680

55

## (2) INFORMATION FOR SEQ ID NO: 367:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1847 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 367:

AAGATTAGGC ATCAAGACAG GGTGCGGATT GTTTGAAATC CCACATAGnA ATGATATTTa 60  
 CaTTATCAAT CCaAGTATGC GTAAATATCT TAATGTTTCA GTTGCTATTT CTAAGATTGC 120  
 ATTGCGTTAT ATTCCACCTG AAGATTTACA CCAATATAGT ATTGACGAAT TTTTATGGA 180  
 TGTACTGAT AGCTATCATA GATTTAGTTC TACAGTACAT GCATTTTGCG AAAGACTTAA 240  
 ACGTGAAATT TATGAAGAAA CAGGCATTTA TTGTACTGTG GGCATTGGTT CTAATATGTT 300  
 ATTAAGTAAA ATTGCTATGG ATGTTGAAGC GAacATAGTc AAAATGGTAT AGCTGAATGG 360  
 CGATATCaAG ATGTACCAAC GAAATTATGG CCAATTcmGC CCtTGCGAGA TTTTGGGGT 420  
 ATTAATCGTC GAACAGAAGC CAAATTGAAT AAAAGAGGAA TTTTACTAT AGGAGATTTA 480  
 GCGAAATATC CATATAAATT TTTAAAAAA GAGTTCGGTA TTTAGGTGT TGATATGCAT 540  
 CTACATGCGA ATGGGATAGA TCAGAGTAAA GTACGTGAAA AGCACAAGAT CAGCAATCCA 600  
 TCGATATGCA AAAGTCAAAT ATTAATGAGA GATTATCATT TTGATGAAGC AAAAGTAGTA 660  
 ATGCAAGAGT TAATTGAAGA TGTTGCTAGC AGAGTTCGAG CAAGAAAAAA AGTGGCAAGA 720  
 ACGATACATT TTGCCTTTGG CTATAGTGAT GAAGGCGGTG TACATAAGCA ATATACTTTG 780  
 AAAGATCCAA CAACTTAGA AAAAGATATT TATAAGTAG TAATGCATTT CGCAGATAAA 840  
 TTATGTAATA AACAAGCACT ATATCGTACG CTAAGTATAT CTTTGAGTCA ATTTATTAAT 900  
 GAGGATGAGC GACgTTAAGT CTGTTTGAAG ATGAATACCA ACGCAAACGT GACGAATGTC 960  
 TAGCTAAAC GATAGACCAA TTACATTTGA AATACGGCAA AGGTATTGTG TCCAAAGCAG 1020  
 TATCGTTTAC AGAAGCAGGT ACAAACACG GCAGATTAGG TTTAATGGCT GGACATAAAA 1080  
 TGTAATGACT ATACGGTTTA AGTAATATAT AACTGTGATT CGTATAAAAT AAGTCTCTAA 1140  
 AGATAAATAT TTCATATATC ACAATAGATT TTCACAATAA TATCTAAGAA TACATGGAAT 1200  
 TTATCAAAAG AGACTTAATA ATTATTGGAT ATAACAATCA AAATCACTCA ATGCTTGCAT 1260  
 ACCGCGTTCT CGGTCAGTAG GGTTTTGA ACTAATTTT AAAGCACCGT ATATATCTTC 1320  
 GCGTACTTCT AAGATTCTTA AGTTGCTTAT AGATATGTTA TGTAACTCA GGATATAAGT 1380

TAGTCCACCT AGTTGTTTAG CGGGTAGTGC GTCGCGATAC GATTTAGCTT GGGCAAAAAA 1500  
 TGATAACAAT TTTTCAGAAT CATTGCTTTC AATTAGTCTT TCTAAATCTT GAAACTGACT 1560  
 5 TTTTAGCTGT CGAATCATTT CTAAAATATA CGTTTTATTA CTCAAGGTGA TATCTTTCCA 1620  
 CATTGTGCA TTACTIONACTAG CTATACGAGT GATATCACGA AAACCACCAG CTGCAAGTTT 1680  
 ATTAATAAAA TGATGTTCTT GACCGTTCTT TTGACTAACA TGAATAAAC TAGATGCAAC 1740  
 10 GATATGAGGT AAATGACTTA CGACGCTTGT TACGTAGTCG TGTTCTTCAG CAGTAGTTAC 1800  
 AATAAATTTA GCAAGAGTAG GTGATAACAG TTCTTTTAAC GTGTTTG 1847

15 (2) INFORMATION FOR SEQ ID NO: 368:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 494 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 368:

25 AACAAAAGGC ATAAGTTCGT GAATTAATGC GTATACAAGG ATAAAGCTTA TAACAGTAGT 60  
 AATGTTGCT ATCAAACGAA CAACATATAT TCTATTTTCA GATAGCAATC GATTCATAA 120  
 TCGATAATTT ACGTATACAA GAATTAACAA CAGCACAATA TAAACAATAA TCATATTTGG 180  
 30 CCCCATTATA TTTTAAATTT TGTTTTTACA TCATTTTCTA CTTTATTATC ATACTAATTT 240  
 TAAAGGCAAA GGTGGACATC GGCAACCTCT CGTAAACTAT TTATCAAAAA TAAACGTATC 300  
 TCATTGTTAT GATATTTATA AATCAATTCG TTTTATTATA AGTCTTTTTT AACAAGTTTG 360  
 35 TCACTATCTA TTAAATAATC ACGCATGCAT CCTTTTAAGA AATCATCTTT ATAACCTGGT 420  
 GTGTACCATT TTCCATCCTC TTCAATGACA ATGTTGCCAA TATCAAATTC AAGGACCTTG 480  
 40 CCGTCCTCTG AAGT 494

(2) INFORMATION FOR SEQ ID NO: 369:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2518 base pairs
  - 45 (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 369:

GTCCATATTT CCCGTCCAAC CAACTAAAAA TGGGTAAATC CCTGGATTTa AATCTACTTC 60

55

	GATAATTTCh	AATGTAGCGA	CCATCATTAC	GAGATGATTT	GATATAAGCA	CAGTTTGGAT	180
	GTGACCAAT	ACTATCGCCT	TCTTCTTCGA	TGATATCTAT	TTTAATACCA	TCATCAGCTG	240
5	CAATTTCTAA	TGAAGATTTA	ATTCGGTTAT	CAAATGTTGA	ATATCCCATT	GCTCCACCCA	300
	CAATAGCGAC	ATCTGTACCA	TGTCCTTGGT	GTGTTTGAGC	AAATGATTCA	TAATAATGTA	360
	TTTCAATATT	TTTAGGAATA	TCTCCCAATA	TTGCGCGTGC	TGAATTCCCA	ATCTTTACTG	420
10	CACCAGCCGT	ATGAGAACTT	GAAGGGCCCA	TCATAACTGG	TCCGATAATA	TCGAAAGCAC	480
	TTTGATAATC	ATAGCTCTTT	GCCATAATTA	AACACTCTCC	TTAATATGAT	TCTTTTGTCT	540
	CGGCATTTTA	AAGTTGATAT	TCATTAAATT	AACTTATTA	ATTAGTGTTT	CAATAATATA	600
15	GGCTAAGACA	ATGCTGACAA	TAATCACTGT	TGCGATTGTT	ACAATTGATG	TCACTGCATT	660
	ATTAAAGCCA	AACAATACGA	TGGCGCCTGC	AATTGGTGT	GCCATACCTT	TGACACCTAT	720
20	TACTAGTCCG	CTAAATGTCA	CGATACATGC	GTGACGACC	CCAATCAGTG	CATTTGTACC	780
	ATATAGTTGT	ACTGGATATT	GCGCTATTAA	ATCAATTTGC	GTCAATGGCT	CAATACAAAC	840
	TGCAAATGCT	TTTGACGGTC	CACCAATGTT	TAATTTTCGG	AATAAAATAA	GGTTAACAAA	900
25	TGAGCTACCT	GTACATGTTA	GTGCTCCAAT	AGCCATAGGA	ACACCTGTCA	GTCCTAATAA	960
	ACTTGTTAAT	ACCAITGAAC	TTAGCGGTGT	CATACCTGTA	ACAGGAATCA	CTAGTCCTAA	1020
	AATGACCGCT	AATGCATATG	GATTGTTATC	ACCTACCGCA	GTGACAGCAC	TACCTATTTG	1080
30	TTTTAATGTT	GCTAGCACAC	CAGGTGTAAT	GATTGATGCA	AGTCCGAAAG	CAATTGCTGG	1140
	TGCAAATAAG	ATCACCACAA	TTAAGTCCAA	GCCTTCTGGA	ACTTTCTTTT	CAATCCATTT	1200
	AATTAAAAAA	GcTACGCCAT	AAGCTGCGAT	GAATGCTGGT	AATAATTTAA	AGTCATGTAA	1260
35	TACTAAACCA	ACAATGACCG	CAAATACTGG	TGCAACGCCT	AAGTTTAAGC	ACGTTAGAAT	1320
	ACCTACTGcG	ATACCGCTTA	AACTTCCTGC	TAAATCCCCA	ATATCTTGTA	GAAATTTAAT	1380
40	ATCAAATACG	CCACCAATAG	CATAACTTAA	GAATGCTTGT	GGTAGAAATG	TCGCACAAGC	1440
	TGCACCGGAT	AATGCTTGTA	GTCCTTGTTT	ACCGTACGGT	GCATACTTTA	AAAATAGCGT	1500
	CATGATCACT	AAAACCAAGA	CTAATGTGCC	TACACCTAAC	AGAATATCCA	TTTCCCCAAA	1560
45	ACCTCTCTCT	ATGTTTATTT	TATTTTCAGA	CCATAAACAT	CGTACACCCT	TAAGAAAACG	1620
	TTTTCAACTT	TTATCTGTTA	TCAAATCaAA	TATTTAAGTG	AAATATTTCA	TATTTGTGAA	1680
	AGATTTTTTAA	AATGGATTGT	TTCAAAAACT	ACTTATGTTG	TCGTTAATAT	TTACTAATTA	1740
50	ACTTTTTACT	CTATATTTCA	AACAGTTGTG	TGACAGTTTT	TTGATAACTT	TTTTACATCT	1800
	GAAAGTAAGT	AATATTTCTA	AAAACTTTTA	ATATTTATAC	ACTTTATCTT	TCGAGCTATT	1860

55

ACACATATAT TTGCAATAAG ATAATTAAAG TAGGATATTA TTTTtagTTT TCTGATAGGA 1980  
 ATGATGATAG TCTATAGGTT GAATCTTTAC TTTTttTTAA AGCTAAATTT ACATCAACTT 2040  
 5 AACAAATGGTT GGTtTACTG AAGATGAAAA TATTTAGTAT AACTTAGTGG AGGCGATAAA 2100  
 GGTGCAATTG AGTCATTCCG TTAAAGTTGC AATTtCTATC TATTTAGCAC TTATCTTTAT 2160  
 AACGTTCACT TCTTATTTAG TCATTATTTT ATATACGAGT ATGACTGGAC ATGATGTATC 2220  
 10 ACATTTCTGT TTAGATAGTC AGCATTCTCA TCATGGATCT CTTACGCAA AACATTTGAG 2280  
 TCTTCTGAA ATCTCATTa AATAGTTTAT CCTCTGTGTT TCAACATTCA TTTCCCATAT 2340  
 CGATTCATTT ATCTATCATC TAGACCACTA CATCTTAGAT GATTTTTTTT TTTCTCATT 2400  
 15 TCACTCTTTC TTAAAGTCG ATATAATGAA TTAAATCATT ATCATAACC GACATATTTT 2460  
 ATGTTGTTGG TGTTAAGTTT aaAGGGGTGA GATACTTGGC GAATaATCaT TCAGCTTT 2518

20 (2) INFORMATION FOR SEQ ID NO: 370:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 790 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 370:  
 30 ATACTACTGG ACCGTChACC AATTGACAAA ACCTGAAAAC TGGATTTTTT TTGCAGGATT 60  
 ACATCGAACC TGGAAGAACA ATCCCCAGGA GTTGATGATA TTCGAGATAT GGCATACAAT 120  
 CAAGGTAGTT TAGATAAGAC AATTTATGAA ATTTCTAAAC GCACAGTACT ATTTTtAATA 180  
 35 CAGAAAGATA TTACGGTATA TAATAAGACG ATTGACTGTT TAAATTATTA TAACTATAGT 240  
 GACGAAAGAA TAAAGGATGA TTAAATGAAT TCACAAGAAT TATTAGCAAT TGCTGTGGAT 300  
 GCAATTGACA ATACCCcAGG CGAAGATACG ATTTCTTTAG AAATGAAAGG TATCAGCGAT 360  
 40 ATGACAGATT ATTTTGTGTT AACGCACGGA AATAATGAAC GACAAGTTCA AGCGATTGCT 420  
 AGAGCGGTGA AAGAAGTAGC CAATGAACAA AATATAGAAG TAAAACGTAT GGAAGGATAC 480  
 AATGAAGCGC GTTGGATATT AATTGACTTA GCTGATGTTG TGGTACATGT TTTCCATAAA 540  
 45 GACGAAAGAA ATTATTATAA TATTGAAAAG TTATATCAAG ATGCACCATT AGAATCATAT 600  
 AGTCAGGTTG CGTATTaATT ATGTCGCAAT ATGCAGAAAT GAGCCTAGTG TACGATCAAT 660  
 50 TGACTCAAGA TCAACCATAT GAAAAATGGT TTGAAATTGT AAAAAATCAC TGCAAAGATG 720  
 AATCAAATAT TTTAGATATT GGATGCGGTA CTGGTAGTTa ACAGTTCAAT TAGAAGCTTT 780

55

## (2) INFORMATION FOR SEQ ID NO: 371:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1823 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 371:

	ATAGATGAAG GTGCAAAATAT TGAAaTAGGT TATTTACCTG GACGCTTGAA ATGGTTAGTT	60
5	GCTGATTTAT TAACTAAACA AGGATTAAAA GTAGTTAACG ACGATATGAC AGGAAGAACG	120
10	TTAAAAGATC GTAAATTATT AACAGGTGAC AGTCCTTTAG CTTCAAATGA GTTAGGAAAA	180
15	TTAGCAGTTA ATGAAATGTT AAATGCAATA CAAAATAAAT AATTAAATAT TAATTAGAGG	240
20	AGCCTCATAT GTAAATGTAT GAGGGCTCTT TTTTGTGGCA AAATTTAAGT GATACTTGTA	300
	AAATAGAACC TATTATGAGT ATGATTTAAG AAAACGCTTG CAAACTAAT AACCGCAACT	360
	AGCGATATGG AGGAAACATG ATGTCTTATA GCATTGGAAT TGATTATGGA ACTGCTTCAG	420
25	GCCGTGTGTT TTAAATTAAT ACAACTAACG GTCAAGTAGT ATCAAAATTT GTGAAACCAT	480
	ATACACATGG TGTCATTGAG AGTGAATTAA ATGGTTTGAA AATACCACAT ACATATGCAC	540
	TTCAAAATAG TAATGATTAT CTAGAAATTA TGAAGAAGG AATATCATAT ATAGTACGTG	600
30	AATCAAAAAT AGATCCAGAC AATATAGTAG GTATTGGTAT AGACTTTACT TCATCTACTA	660
	TTATTTTAC TGACGAAAAC CTTAACCCGG TACATAACTT AAAACAATTT AAAACAATC	720
	CACATGCGTA TGTGAAACTT TGGAAACATC ATGGTGCATA TAAAGAGGCA GAGAAATTAT	780
35	ATCAAACTGC TATTGAAAAT AATAATAAGT GGTTAGGCCA TTATGGATAT AATGTTAGTA	840
	GTGAATGGAT GATTCCCAAA ATAATGGAGG TCATGAATCG AGCACCAGAA ATTATGGAAA	900
40	AAACGGCTTA TATTATGGAA GCGGGCGATT GGATTGTAAA TAAATTAACT AATAAAAATG	960
	TACGCTCGAA TTGTGGATTA GGTTCCAAAG CATTTTGGGA AGAAGAAACA GGGTTTCATT	1020
	ATGATTTATT TGATAAAATA GACCCCAAAT TATCAAAAGT AATTCAAGAT AAAGTATCTG	1080
45	CACCGTTGT TAATATTGGT GAAGCAGTAG GGAACTGGA TGATAAAATG GCACAGAAAT	1140
	TAGGATTATC AAAAGAAACT ATGGTAAGTC CTTTATTAT TGATGCCCAT GCTAGTTTAT	1200
	TAGGTATTGG GTCTGAAAAA GATAAAGAAA TGAATATGGT GATGGGAACA AGCACATGCC	1260
50	ATCTTATGTT AAATGAAAAG CAACATCAAG TGCCAGGTAT ATCAGGTTCT GTAAAAGGAG	1320
	CAATTATTC AGAATTATTT GCTTATGAAG CGGGGCAATC AGCAGTAGGT GATTGTTTG	1380

55

CTGTATTTGA ATTAATGAAT GAAAAGATAA AACATCAAAT GCCAGGTGAA AGTGGGCTCA 1500  
 TTGCTCTTGA TTGGCATAAT GGAAATCGAA GTGTATTAAG TGATAGCAAT TTAACAGGTT 1560  
 5 GTATCTTTGG ATTAACTTTA CAAACTAAGC ATGAGGaTAT TTATAGAGCm TATTTaGaAG 1620  
 CTACaGCATT TGGTACTAAG ATGATTATGC mACAGTATCA AGATTGGCAT ATGGaAGTAG 1680  
 aAAAGGtATT TGCaTGTGGc gGTATACcTA AAAAGAATGC TGTTATGATG GATATCTATG 1740  
 10 CGAATGTACT GAATAAAAAA CTAATTGTTA TGGATAGTGA GTATGCACCA GCAATAGGCG 1800  
 CAGCAATATT AGGTGCAGTC AGT 1823

## (2) INFORMATION FOR SEQ ID NO: 372:

15 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1600 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 372:

25 ACGATCATCT GCATCAGCGA ATTCCGATGc AaTtCaTATG tCcTaATAAC AAAAGAATAG 60  
 GTTTTaAAAAG ATATGCATAC CGTAATGaTA GATATAGTTT TAAACGTGAC TTCAAGCTAT 120  
 ATGaATGTGA TGA CTGTTCA TCATGTTCTT TGAGACATCA ATGCATGAAG CCAaATTCGA 180  
 30 AATCCAATAA GAAAATTATG AAGAATTATA ATTGGGAATA CTTTAAAGCC CAAATTAATC 240  
 AAAAGCTTTC TGAACCAGAA ACGAAAAAAT CTATAGTCAA AGAAAAATTG ATGTAGAGCC 300  
 TGTTTTTGGA TTTATGAAGG CTATTTTGGG TtTCACTCGA ATGTCAGTTC GGAATAAAT 360  
 35 AAAGTTAAAC GAGAGCTAGG TTTTGTATTA ATGGCACTTA ATATAAGGAA AATAgCaGCT 420  
 CAACGAGCTG TACATTATAA AATACATATC AAAAAAGCTG ATTTCTATCA AATAAATAAT 480  
 AGAAATCAGC TTTTTTACAT TGCCTAAGAA CTTTAAGGAA CTTAATGTCC CAAGCTCTTT 540  
 40 TTTGTTATAT CTAATTCGTA ATTTATGATT GTTTATTCGG TCCTTTGATG TTTCACTAAAT 600  
 GTGACTTTAA ATCTTGTTCT AATTGTTGTA ATTCTTTTTC AGCTAATTGT CGTTCTTCGC 660  
 45 GACCGTGTG TTGAATAATT AATGTTTCTT CAATTGTCTC AATAATGTTA CGCTGTGTAC 720  
 GTTCAATGT ATCAAGATCA ACAATGCCAC GCTCATTTTC TGTTGCAGTT TCAATCGCAT 780  
 TTTGTTTCAA CATTTAGCA TTTGCTGTTA ATAAATCATT AGTTGTATCA GTGACAGCTC 840  
 50 GTTGTGCAGC AACTGCATTA CGCTGTCTCA TTAATGTAAG CGCAATGGCC ATTTGATTTT 900  
 TCCATAGTGG AATACTTGTC AAAATTGAAC TTTGTATCTT CTCGGCAAGT GCTTGATTAA 960

55



AGTCATATAT GCGTTTATCT AGTCTATCTA TAAATTGCTG CATATCTGCA ACTTGTTGAA 1080  
 TATCCATTG ATTAGTGGAT TgtGCGCTTG CTGTTGCAAT TGTGGTAGCT TTTCATTTTC 1140  
 5 TAATTGCAAC TTTTTTTGcT GTGCAGCAAT GATATGCAAT GATAAGTCAT CAAAGTATTG 1200  
 TTTGTTTTTA TCATATAGCG TATCTAATAA TTCAATATCT CTTGTTAAAT GTGTTTGATG 1260  
 TTTCTGCAGT TGAATCGTTA TGCGATCGAC TTGAGCACTA ACTGATTGCA TTCTTGAAAA 1320  
 10 GATTTTCATTG ATAGACGACT TTGCTCTGCT AAAAATTCTT TTAAACATAG ATGGTTTATC 1380  
 AGTATTTAAC TCATTGGAT TAACTGACTT TAGTTTTGAC ATTAGATCTG ACAAAGTATC 1440  
 15 TCCAATAGGA CCAACATCTT TACTTTGTAC TTCATCCAAC ATTTGATGTG AAAATTGAGA 1500  
 CATTTGTTTC TGkAAATCAG AACCAAACGC TAATAAACCT TCATTGTCTA AAGGGTTAAT 1560  
 TTGTTTACTG ATTGTGTCTA CCTGTTTTTG TTGTTCAATT 1600

20 (2) INFORMATION FOR SEQ ID NO: 373:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1227 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 373:

30 ATGTTGATAA TGGTAGCTTC TACAAGAATA AAGACCAACA AGTTGGTGCA ACAATTCTTG 60  
 aTAGTAAAC TGGTGGTTTA GTTGCTATAT CTGGTGGACG TGATTTCAAA GACGTCGTTA 120  
 ACAGAAACCA AGCAACAGAT CCTCACCCTA CTGGTTCATC TTAAAACCT TTCTTAGCGT 180  
 35 ATGGACCTGC CATTGAAAAT ATGAAATGGG CAACAAACCA TGCGATTCAA GATGAATCTT 240  
 CATATCAAGT TGATGGTTCT ACATTTAGAA ACTATGATAC GAAGAGTCAC GGTACTGTAT 300  
 CTATTTATGA TGCTTTACGA CAAAGTTTCA ATATCCCAGC TTAAAAGCT TGGCAATCAG 360  
 40 TTAAGCAAAA TGCTGGTAAT GATGCACCTA AGAAATTCGC TGCCAAACTT GGCTTAACT 420  
 ACGAAGGCGA TATTGGTCCA TCTGAAGTAC TTGGTGGTTC TGCTTCAGAA TTCTACCAA 480  
 45 CACAATTAGC ATCAGCATTT GCTGCAATCG CTAACGGTGG TACTTATAAC AACGCGCATT 540  
 CAATTCAAAA AGTAGTTACT CGTGATGGTG AAACAATCGA ATACGATCAT ACTAGCCATA 600  
 AAGCGATGAG TGATTACACT GCATACATGT TAGCTGAGAT GCTAAAAGGT ACATTTAAAC 660  
 50 CATATGGTTC TGCATATGGC CATGGTGTAT CTGGAGTAAA TATGGGTGCT AAGACAGGTA 720  
 CTGGTACTTA CGGTGCTGAA ACTTATTCAC AATATAATTT ACCTGATAAT GCAGCGAAAG 780

55

AAGTTAAACA ATATGGTGAA AACTCATTTG TrGGACATAG CCAACAAGAA TATCCACAGT 900  
 TCTTATATGA AAATGTGATG TCAAAAATTT CATCTAGAGA TGGCGAAGAC TTAAACGTC 960  
 5 CTAGCTCAGT AAGTGGTAGT ATCCCATCAA TCAATGTTTC TGGTAGTCAA GATAACAACA 1020  
 CTACAAATCG TAGTACACAC GGTGGTAGTG ACACATCAGC AAACAGCAGT GGTACTGCAC 1080  
 AATCAAATAA CAATACTAGA TCTCAACAAT CTAGAAACAG CGGTGGATTA ACAGGTATAT 1140  
 10 TCAACTAATC CACTCAACAT AAAATCCTCA GTTATACCAT ATTTATGGTG TAGCCGAGGA 1200  
 TTTTnTTAGG TTCTTCATCT TTTATGG 1227

(2) INFORMATION FOR SEQ ID NO: 374:  
 15

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1953 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 374:

25 CCATATGGtG CAACATTCTT CGTATTTAGT GATTATTTAA AACCAGCGTT ACGTTTATCA 60  
 TCAATTATGG GATTAAAtgC aACGTTcATC TTCACACATG AyTcaATTGC AGTAGGTGAA 120  
 GATGGTCCTA CTCATGAACC AATTGAGCAA TTAGCTGGAT TAAGAGCCAT TCCAAATATG 180  
 30 AATGTTATCC GTCCTGCTGA TGGTAATGAA ACAAGAGTAG CATGGGAAGT TGCCTTAGAA 240  
 TCTGAATCTA CACCTACTTC ATTAGTATTG ACACGTCAAA ACTTACCGGT ATTAGATGTA 300  
 CCAGAAGATG TAGTTGAAGA AGGCGTTCGA AAAGTGCCCT ATACAGTTTA TGGCTCTGAA 360  
 35 GAGACACCAG AATTCCTATT ATTAGCTTCA GGTTCAGAAG TTAGTCTTGC AGTTGAAGCT 420  
 GCTAAAGATC TTGAAAAACA AGGTAAATCA GTACGTGTTG TTTCAATGCC TAACTGGAAT 480  
 GCATTTGAAC AACAATCTGA AGAATATAAA GAATCAGTTA TTCCATCAAG CGTAACAAAA 540  
 40 CGTGTTGCGA TTGAAATGGC TTCACCGCTT GGATGGCATA AATATGTAGG TACTGCAGGT 600  
 AAAGTTATTG CTATTGACGG CTTTGGCGCA AGTGCACCTG GCGATTTAGT AGTTGAAAAA 660  
 TATGGATTTA CAAAAGAAAA TATCTTAAAC CAAGTTATGA GCTTATAAGA ATAATTTATA 720  
 45 AAGCGAGTAT GTTTAGAAGT CTAGGATGCA TAATCTTAGG CTTCTTTTAA AGTGTGAAA 780  
 TTTAGAGTAT AGCACTTAAA CTACATCATA AGTGATAAGT TATGAAAGTA TACTATTTCA 840  
 50 GATTAATCTT TAAAAGCTCT GTTATAACAG CATGATTTTT GATATTATTT TTAGTATCGA 900  
 TATTAAAATA CTTGAATAAA CTAGTTCTTG AAATAATGTG ATGAATTTAG TAAAATTCAG 960

55

TGCATTAATT TTAGGTTTAA TTGGAGGTTT CCTTTTAGCT AGAAAATATA TGATGGACTA 1080  
 CTTGAAGAAA AACCCACCAA TCAACGAAGA AATGCTTCGT ATGATGATGA TGCAAATGGG 1140  
 5 TCAAAAACCT TCTCAGAAGA AAATTAATCA AATGATGACG ATGATGAATA AAAATATGGA 1200  
 TCAAAATATG AAGAGTGC GA AAAAGTAAAT TCGCAATTGA TAGAGGCTAT TTTCCAGATA 1260  
 TGGAAATGGC CTCTTTTAT AATCAAATTA ATAAGAATA ATATGTTTAT TAAAATTAAA 1320  
 10 GTTAACAAA TGACGAATAG ACTGAGAAAT GCTATAATTC ATTTTGTATG ATTTACAGAG 1380  
 AGTTTATTTA ACGAGAAGGT GTCyGCGTGC TCTATTTAAT ATTTTCAATC ATTGTAGCTT 1440  
 TATTTATGGG AACTATAGTT ATAGTTATTC GTATGAAAGC TCAAAATTAT CCGGTAAaKG 1500  
 15 AGAAAAAAT AGTTTGGCa CCGTTTTTTA tGGCgACCGG TGCATTGATG TACGTCGTTC 1560  
 CaTATTTTAG GCTAACAGGA TCGGAAATGC TAGAAGCCTT TATAATTGGT TTGCTTTTTT 1620  
 CtACAGTTCT AATTGGACT TCTCGATTG AAGTCAAAGG TACAGAAATT TATATGAAAC 1680  
 20 GATCTAAAGC ATTTCCAGTT ATTTGATTT CATTACTTAT CATTCTACT GTGATGAAAA 1740  
 TATTCATTAG TAATGAAATA GATCCTGGAG AATTAGGCGG CATGTTCTTT TTATTAGCAT 1800  
 25 TCTGTATGAT TGTTCTTGG AGAGCAGCAA TGCTATATAA ATACAAAAA CTAAAGAAAA 1860  
 CATTAATCAA TTAATTACTT TTaAAACCAC TTGTGATCGA CTTCTAAATC AGTCAATGAG 1920  
 TGGGTTTAAT nTTACTTGGA AAAGGnGGAA AGG 1953

30 (2) INFORMATION FOR SEQ ID NO: 375:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3787 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 375:

40 ACATTTGATC AAnTTATCGA CATTAAAGAT GAATTCAnTT GATCGTnTCA ATGATTATCC 60  
 TGTTGAAGTA GCACGTTTGC TTGATATAGT GCAmATaAAA GTACACGCAT TACATTCAGG 120  
 45 TATCcACGTT GATTAAAGAT AAAGGGAAAA TAATTGATAT TCATTTATCT GTAAAGCCA 180  
 CTGAAATAT TGATGGCGAA GTGCTGTTCA AAGCAACACA ACCTTTAGGT AGAACAATGA 240  
 AGGTTGGTGT TCAAAATAAT GCAATGrCAA TTACTTTAAC GAAACAAAAT CAATGGCTTG 300  
 50 ATAGTTTGAA GTTTTAGTT AAGTGCATTG AAGAAAGTAT GAGAATCAGT GATGAAGCAT 360  
 AAAGAAGCAT TTAATGGCGT TGTCGTGTTA ACTGCTGCAT TAATTGTCAT TAAAATTCTG 420

55

CAACAAGTGT ATCCAATTGT AGCATTAGGA ATGATATTAT CGATGAATGC CATTCCCTAGT 540  
 GCAATTACAC AAAATATAGG GAAGTATCAT AGTGACGAAG CATATGCAAA AGCAGTCGCT 600  
 5 TATATACAAT TAGTTGGTAT ATTATTATTT ATTGCTATTT TTGTGTTTGC GAACAATATT 660  
 GCACATATGA TGGGTGATGG CCATTTAACA CCAATGATTC AAGCTGCAAG TTTAAGCTTT 720  
 ATATTTATAG GTATGCTTGG CGTGTTAAGA GGTATTATTC AATCTGCAAA TAATATGACA 780  
 10 GTTCCGGCTA TTTCCAGGT TATAGAACAA GTTATACGAG TAGGTATTAT CATTGTTACT 840  
 ATTGTTATTT TTGTAGACAG AGGTTGGACG ATATATGAAG CGGGAACAAT TGCTATTTTA 900  
 GCATCAACGA TAGGTTTTTT AGGTTCTTCA ATTTATTTAG TAGCGCACCG ACCTTTTAAG 960  
 15 TTTAAAATGG TAAATAACAC TGCAAAGATC GTTTGGAAAC AGTTCGCACT TTCGGTTTTG 1020  
 ATTTTCGCTA TCAGTCAATT AATCGTAATT TTATGGCAAG TGATTGATAG TGTTACTATT 1080  
 ATTAAGTCAC TTCAAGCGAT ACGCGTGCCA TTCGATGTTG CCATAACTGA AAAAGGAGTC 1140  
 20 TATGACCGTG GTGCATCATT TATTCAGATG GGATTGATTG TAACTACAAC ATTTAGTTTT 1200  
 GCGCTCATT CTCTGTTAAG TGACGCAATC AAAATGAATA ATCAGGTACT TATGAATCGT 1260  
 25 TATGCAAATG CGTCATTAAA GATTACGATT TTAATAAGTA CAGCAGCGGG AATAGGATTA 1320  
 ATTAATTTAT TGCCTTTAAT GAACGGTGTG TTTTTAAGA CGAATGATTT AACCTTAACG 1380  
 TTAAGTGTAT ATATGATTAC GGTCAATTGT GTATCGTTAA TTATGATGGA TATGGCATT 1440  
 30 TTACAAGCGC AACATGCTGT GAGACCTATT TTTGTGGA TGACGGCAGG ATTGGTTATT 1500  
 AAATTATAC TTAATATCAT TTTGATTCGT TTAAGTGGA TTATTGGTGC GAGCATTAGT 1560  
 ACTGTGTAT CATTAAATTAT ATTCGGTACG ATTATCCATA TTGCTGTCAC GAGAAAATAC 1620  
 35 CACTTATATG CGATGAGACG ATTTTTTATC AATGTTGTTT TAGGTATGGT ATTTATGTCC 1680  
 ATTGTTGTTT AATGCGTGT AAACATAGTG ACAACACACG GTAGAATCAC TGGACTCATT 1740  
 GAATTATTAT GTGCAGCAGT ATTAGGTATC ATTGCATTGT TTTTCTATAT TTTTAGATTT 1800  
 40 AATGTTTTGA CATATAAAGA GTTAACTTAT TTACCATTGT GTTCAAAGTT GTATCAAATT 1860  
 AAGAAAGGAA GACGTTGATG GCACATACCA TTACGATTGT TGGCTTAGGA AACTATGGCA 1920  
 TTGATGATTT GCCGCTAGGG ATATATAAAT TTTTAAAGAC ACAAGATAAA GTTTATGCAA 1980  
 45 GAACGTTAGA TCATCCAGTT ATAGAATCAT TGCAAGATGA ATTAACATTT CAGAGTTTTG 2040  
 ACCATGTTTA TGAAGCACAT AACCAATTTG AAGATGTCTA TATTGATATT GTGGCGCAAT 2100  
 50 TGGTTGAAGC TGCTAATGAA AAAGATATTG TCTATGCGGT TCCGGGTCAT CCTAGAGTTG 2160  
 CTGAGACAAC TACAGTGAAG TTAGTGCTT TAGCAAAGGA CAATACTGAT ATAGATGTGA 2220

55

	ATGATGGCTT CACACTGTTA GATGCGACAT CATTACAAGA AGTAACACTT AATGTTAGAA	2340
	CGCATACATT GATTACGCAA GTTTATAGTG CAATGGTTGC TGCTAATTG AAAATCACTT	2400
5	TAATGGAACG ATATCCTGAT GATTACCCTG TTCAAATTGT CACTGGTGCA CGAAGCGATG	2460
	GTGCGGATAA CGTTGTGACA TGCCCATAT ATGAATTGGA TCATGATGAA AATGCATTCA	2520
	ATAATTTGAC GAGTGTATTC GTACCAAAAA TCATAACATC GACATATTTG TATCATGACT	2580
10	TTGATTTTGC AACGGAAGTG ATTGATACTT TAGTTGATGA AGATAAAGGT TGTCCATGGG	2640
	ATAAAGTGCA AACGCaTGma AcgCTAAAGC GTTATTTACT TGAAGAAACA TTTGAATTGT	2700
	TCGAAGCTAT TGACAATGAA GATGATTGGC ATATGATTGA AGAACTAGGA GATATTTTAT	2760
15	TACAAGTGT ATTGCATACT AGTATTGGTA AAAAAGAAGG GTATATCGAC ATTAAAGAAG	2820
	TGATTACAAG TCTTAATGCT AAAATGATTC GTAGACACCC ACACATATTT GGTGATGCCA	2880
20	ATGCTGAAAC TATCGATGAC TTAAGAGAAA TTTGGTCTAA GCGGAAAGAT GCTGAAGGTA	2940
	AACAGCCAAG AGTTAAATTT GAAAAAGTAT TTGCAGAGCA TTTTTTAAAT TTATATGAGA	3000
	AGACGAAGGA TAAGTCATTT GATGAGGCCG CGTTAAAGCA GTGGCTAGAA AAAGGGGAGA	3060
25	GTAATACATG AGATTAGATA AATATTTAAA AGTATCACGG TTAATAAAGC GACGTACGCT	3120
	AGCAAAAGAA GTAAGTGATC AAGGTAGAAT TACAATAAAT GGTAATGTTG CTAAAGCTGG	3180
	ATCGGATGTT AAAGTTGAAG ATGTGCTGAC GATTCGCTTT GGTCAAAAAT TAGTAACAGT	3240
30	TAAAGTAACT GCATTAAATG AACATGCATC TAAAGATAAC GCGAAGGGTA TGTATGAAAT	3300
	CATTGAAGAG CGTCGACTTG AAGAAGCGTA AATTGGAGGT GACAAGCAAT GAAAAATAAA	3360
	GTAGAACATA TAGAAAATCA GTACACGTCG CAAGAGAACA AGAAAAACA ACGTCAAAAA	3420
35	ATGAAAATGC GTGTTGTTc TAGGCGTATT ACAGTATTTG cGGGCGTATT aCTTGCATA	3480
	ATTGTTGTTT TATCaATCTT GCTTGTGTG CAAAAACATC GCAATGATAT TGATGCACAG	3540
	GAGCGAAAAG CGAAAGAAGC ACAGTTTCAA AAGCAACAAA ATGAAGAAAT TGCCTTAAAA	3600
40	GAAAAGTTGA ATAATCTGAA TGACAAAGAT TACATTGAAA AAATTGCGCG TGATGATTAT	3660
	TACTTAAGCA ACAAAGGTGA AGTGATTTTT AGGTTGCCAG AAGACAAAGA TTCGTCTAGC	3720
45	TCAAAATCTT CGAAAAAATA AATCCAAATT GATTCAAAAT TATCCGAGTA TAGACATTGT	3780
	GAAAAAA	3787

(2) INFORMATION FOR SEQ ID NO: 376:

- 50 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1644 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 376:

5	TAAACCATTT CAAACTGAGG AACGCnAAGA CGGsACgTTT CCAGATTTAG AAGTATTTAA	60
	AAATGAATGT GATTTAAGCT ATGACATAAC GTCACTTTAT ACTTTTAAGC AACCTGTATC	120
	ACCACACCTT GCATTTAAAA TGACAGATCA AATTTTCTA AATAAGCAGC GTGTATTAGA	180
10	TAAGGTAAAA GTTTTAGATA AGGAATTTGA TTTTATCTTA ATTGAGGGTG CTGGGGGAAT	240
	TGCCGTACCA ATATATGAAG GTACAGATGA TTTCTACATG ACTAAAGATC TAATCAATGA	300
	TTGTGCAGAT TGTGTCATCA GTGTGTTGCC ATCAAAATTA GGTGCTATTA GCGATGCCAT	360
15	TGTTCCACCA GATTATGTTA ATCAGAATGT ATCGGCGAGT AATTTTTTAA TAATGAATCG	420
	CTATACAGAC AGCTATATTG AAAAAGACAA TCAATGACG ATTGGAAAAT TAACAAATAA	480
	AACAGTCTAT ACATTTGAAG AACATGCCAC GTATGAAAAT TTCTCAGAAG CATTTTTTAA	540
20	ACAATTAATA GGAGTTAAAA ATGAATTACA CACAACAACT TAAACAAAAA GACTCAGAAT	600
	ATGTTTGGCA TCCATTTACA CAAATGGGTG TATATAGCAA AGAAGAAGCA ATCATCATTG	660
	AAAAAGGAAA GGGTAGTTAC CTTTACGATA CGAATGGCaA TAAATATTTA GATGGTTATG	720
25	CATCGTTGTG GGTCAATGTG CATGGTCATA ATAACAAATA CTTnAATAAG GTAATTAAAA	780
	AGCAACTCAA TAAAATTGCC CATTTCTACGC TGCTAGGATC ATCAAATATT CCGTCAATAG	840
	AACCTGCGGA AAAATTAATC GAAATCACGC CAAGTAATCT AAGAAAAGTA TTTTATTCTG	900
30	ATACAGGCAG TCGTCTGTT GAAATCGCAA TAAAGATGGC ATATCAGTAT TGGAAAAATA	960
	TTGATAGAGA AAAATATGCC AAGAAAAACA AGTTTATAAC GCTAAATCAC GGTTATCATG	1020
35	GGGATACGAT TGGTGCGGTA AGTGTGGTG GTATCAAGAC CTTTCATAAA ATATTTAAAG	1080
	ACTTAATATT TGAGAATATT CAAGTAGAAA GCCCATCTTT CTATCGCAGT AATTACGATA	1140
	CTGAAAATGA AATGATGACA GCTATTTTAA CGAATATAGA GCAAATTCTA ATTGAAAGAA	1200
40	ATGATGAAAT CGCAGGGTTT ATATTGGAAC CGTTGATTCA AGGTGCGACA GGCTTGTTTG	1260
	TTCATCCTAA AGGCTTTTTG AAAGAAGTCG AGAAATTGTG CAAAAAATAC GATGTCTTAT	1320
	TAATTTGTGA TGAGGTAGCA GTTGGTTTTG GGAGAACTGG AAAGATGTTT GCATGCAATC	1380
45	ATGAAGATGT TCAACCGGAT ATTATGTGTT TAGGTAAGGC GATTACTGGT GGCTACTTAC	1440
	CACTTGCAGC TACATTGACA TCTAAAAAAA TATACAATGC ATTTTTAAGT GATTTCGCATG	1500
50	GTGTGAATAC CTTTTTCCAT GGTCaTACAT ACaCCGAAA TCAAAcGTT TGTaCGGTTG	1560
	cATTaGaAAA TATaAGaCTT TATGaAAAAc GTaAGTtnAT TGTgCACATa TTGaAACGaC	1620

55

## (2) INFORMATION FOR SEQ ID NO: 377:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 431 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 377:

ATGATTTTtA aAAATCATTa AGTTAAGGT<sub>r</sub> GATACACATC TTGTCATATG ATCAAATGGT 60  
 TTCGCCAAAA ATCAATAATC AGACAACAAA ATGTGCGAAC TCGATATTTT ACACGACTCT 120  
 CTTTACCAAT TCTGCCCCGA ATTACACTTA AAACGACTCA ACAGCTTAAC GTTGGCTTGC 180  
 CACGcmTTAC TTGACTGTAA AACTCTCACT CTTACCGAAC TTGGCCGTAA CCTGCCAACC 240  
 AAAGCGAGAA CAAACATAA CATCAAACGA ATCGACCGAT TGTTAGGTAA TCGTCACCTC 300  
 CACAAAGAGC GACTCGCTGT ATACCGTTGG CATGCTAGCT TTATCTGTTC GGGCAATACG 360  
 aTGCCCATTG TACTTGTtGA CTGGTCTGat ATCcGTGAGC AAAACGGCT TATGGTnTTG 420  
 CGAGCTTCAG T 431

## (2) INFORMATION FOR SEQ ID NO: 378:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2006 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 378:

TTTnTATAAC GTATTATAAA TCGTTAAAAA TTTTGGTTGT GTTTGCGTCA CGTAGACAAC 60  
 CTCCATAAAG TTA<sub>r</sub>CTTAATC ACTCTCATCA TACAATAATT TTTACTCAA<sub>r</sub> TTGGAAnAAT 120  
 TATAAAAATT AAATATAGAT AGGCTTTGAA AATTAGTTTT ATACAAGGTT AGTAGCTGTA 180  
 ACTGTAAAAT GTTCTTAATA TTGTCAAAAT GTAATGCTTG AAAGCGCTTT TAAaAATAT 240  
 TATTATATAC ATGGTTAGAC AAATAGACAA ATCACTATAC AAATATTGGG AGGAATATTT 300  
 TATGAAATCA ACACCACACA TTAAACCAAT GAATGACGTC GAAATTGCAG AAACGGTTCT 360  
 ATTGCCAGGA GATCCGTTAA GAGCTAAGTT CATTGCAGAA ACTTATTTGG ATGATGTGGA 420  
 ACAGTTCAAT ACAGTGCGAA ACATGTTTGG TTTTACCGGA ACATATAAAG GTAAAAAAGT 480  
 TTCTGTATG GGTTCAGGTA TGkGTATGCC ATCTATTGGC ATTTACTCTT ATGAATTAAT 540

	CATTGATTTA TATGATGTGA TTaTTkCACA AGGTGCCTCT ACTGATTCAA ATTACGTTCA	660
	ACAATATCAA TTACCAGGTC ATTTTGCGCC AATTGCTTCT TATCAATTAT TAGAAAAAGC	720
5	AGTTGAAACA GCACGTGACA AAGGTGTACG TCATCATGTA GGTAATGTGT TATCAAGTGA	780
	TATTTTCTAT AACCGGGATA CAACAGCGAG TGAACGTTGG ATGCGTATGG GTATTTTAGG	840
	TGTAGAAATG GAATCaGCTG CaTTATACAT GaATGCAaTT TACGCTGGTG TCGAAGCATT	900
10	AGGTGTGTTC ACAGTGAGCG ATCATTTAAT TCATGAAACG TCAACAACAC CTGAGGAAAG	960
	GGAACGTGCA TTTaCAGATA TGATTGAAAT TGCCTGTCA TTGGTGTAGA TGATTATGAA	1020
	TGTTGAATAT TCTAAAATAA AGAAAGCAGT ACCTATTTTA TTATTCTTAT TTGTATTCAG	1080
15	TTTGTTTATA GACAACTCAT TTAAATTGAT TTCTGTAGCC ATTGCTGATG ACTTAAACAT	1140
	ATCTGTAACG ACAGTAAGTT GGCAAGCGAC ATTAGCCGGT TTAGTAATTG GTATTGGCGC	1200
	TGTAGTATAC GCTTCATTAT CTGATGCCAT TAGTATACGC ACACTATTTA TTTATGGCGT	1260
20	GATATTAATC ATTATCGGAT CAATTATTGG TTACATTTTC CAACATCAAT TCCCATTACT	1320
	TTTAGTTGGA CGTATTATTC AAACTGCCGG TTTAGCTGCT GCAGAGACAT TATATGTGAT	1380
25	ATATGTTGCA AAGTATCTTT CTAAAGAGGA CCAGAAGACT TACCTTGGCT TAAGTACGAG	1440
	CAGTTATTCC TTGTCATTAG TTATCGGTAC ATTATCAGGT GGATTTATTT CTACGTATTT	1500
	ACACTGGACA AATATGTTTT TAATTGCATT AATCGTAGTA TTTACGTTGC CATTCTTATT	1560
30	TAAATTATTa CCAAAAGAAA ATAATACGAA TAAAGCTCAT TTAGATTTTG TTGGCTTAAT	1620
	TCTAGTGGCA ACTATTGCTA CAACAGTCAT GCTGTTTATT ACGAACTTTA ATTGGTTATA	1680
	TATGATTGGT GCCTTAATTG CGATTATCGT TTTTGCGCTA TATATTAAAA ATGCGCAACG	1740
35	TCCATTAGTA AATAAATCAT TTTTCCAAAA TAAACGTTAT GCTTCATTTT TATTTATAGT	1800
	ATTTGTAATG TATGCTATCC AATTGGGTTA TATTTTACG TTCCCATTCA TAATGGAGCA	1860
	AATTTATCAT CTGcAACTAG ACACAACATC ACTGTTATTA GTACCGGGgT TaTATAGTAG	1920
40	CAGTCATTGT TGGtGgCACT AAGTGGgTtA AAATCGGgCG rAATATCTGG AATTCCAAAA	1980
	CCAAGCGGAT TATCACAGCC AATTAA	2006

(2) INFORMATION FOR SEQ ID NO: 379:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4858 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear
- 55



	TGGGGAAAAA AAGACCCAGC AGTATTAGAA GAATCGTTAA ATATTTCTAT TGAAGAAATG	60
	AATCGTATCA TAAATTAGT CGAAGAATTA CTTGAATTGA CTAAAGGAGA TGTAATGAC	120
5	ATTTCTTCTG AAGCACAGAC CGTGCAATTT AATGATGAAA TTCGCTCGCG AATACACTCA	180
	TTAAAACAAT TGCATCCTGA TTATCAATTT GATACGGATC TGACATCTAA AAATCTAGAA	240
10	ATTAAATGA AACCTCATCA ATTCGAACAA TTATTTTTTAA TCTTTATTGA TAATGCAATC	300
	AAATATGATG TGAAGAATAA GAAAATTAAA GTTAAGACAA GGTAAAAAA TAAGCAAAAA	360
	ATAATTGAAA TTACAGATCA TGAATTGGT ATTCCAGAGG AAGATCAAGA TTTCATTTTTT	420
15	GATCGCTTTT ATCgAGTGGA TAAATCTCGT TCAAGAAGTC AAGGCGGTAA TGGACTCGGA	480
	TTATCTATTG CTCAAAAAAT CATTCAATTA AACGGAGGAT CGATTAAAAAT TAAAAGTGAA	540
	ATTAACAAAG GAACAACGTT TAAAATCATA TTTTAATCAT GACTGAGACG TCAATCAAAG	600
20	TCATAGGATC AATTTTTTAA GTACACATTA GCTGTGACTA ATGTATAAGA ACAACTATAA	660
	AACAAATAAA CAGTGGTTCT TTATCATTTT TGTTGTACTC CCAAATTTA CAATAAAATA	720
	CATCTATAAA CCTAGAAGAA TCAACGCTTT TGTTGATTCT TCTTTTTAGC AGATAAATAG	780
25	GTAAATCTAC TTTAACAAT AACTAAATAG TGATATTATT ACATTGTAAG CGTTTCAACA	840
	TTTTTGTGGA GGGTGTA AAA TGACTAACGA AAGAAAAGAA GTTTCAGAGG CTCCTGTAAA	900
30	CTTCGGTGCG AATTTAGGTC TAATGTTAGA TCTATATGAT GACTTTTTTAC AAGATCCATC	960
	ATCTGTACCA GAAGATTTAC AAGTCTTATT CAGCACAAAT AAGAATGATG ACTCAATTGT	1020
	ACCAGCTTTA AAAAGTACAA GTAGTCAAAA TAGCGACGGC ACAATTAAGC GTGTCATGCG	1080
35	TTTAATTGAT AATATTCGCC AATACGGGCA TCTTAAAGCC GATATTTATC CTGTAAATCC	1140
	TCCAAAAAGG AAACATGTAC CTAAATTAGA GATTGAAGAC TTTGATTTAG ATCAACAGAC	1200
	TTTGGAAGGT ATATCAGCAG GAATTGTTTC AGATCACTTT GCCGACATTT ATGATAATGC	1260
40	TTATGAAGCA ATTTTAAGAA TGGAAAAACG TTACAAAGGA CCAATTGCAT TTGAGTATAC	1320
	ACATATTAAT AACAATACCG AACGTGGTTG GTTAAAAAGA AGAATTGAAA CGCCATATAA	1380
	AGTAACGTTA AATAATAACG AAAAAAGGGC ACTATTCAA CAATTAGCGT ATGTTGAAGG	1440
45	GTTTGAAAAA TATCTTCATA AAACTTCGT TGGTGCAAAG CGTTTTTCAA TTGAAGGGGT	1500
	AGACGCACTT GTACCGATGT TACAACGTAC TATTACGATT GCTGCGAAAG AAGGTATTAA	1560
50	AAATATACAA ATAGGCATGG CTCACCGTGG ACGTTTAAAC GTTTTAACGC ATGTCTTAGA	1620
	AAAACCGTAC GAAATGATGA TTTCAGAATT TATGCATACA GATCCAATGA AATTCTTACC	1680
	TGAAGATGGT AGCTTGCACT TAACTGCTGG ATGGAAGTGGT GATGTGAAAT ATCACCTTGG	1740
55		

	AAGTCACTTG	GAAATTGTTG	CACCTGTTGT	TGAGGGGCGT	ACGAGAGCAG	CACAAGATGA	1860
	TACACAACGA	GCTGGGGCTC	CGACGACTGA	TCATCATAAA	GCAATGCCAA	TTATTATACA	1920
5	TGGCGATGCT	GCTTATCCTG	GTCAAGGAAT	TAACCTCGAA	ACAATGAACT	TAGGAAACTT	1980
	GAAAGGCTAT	TCTACGGGTG	GTTCAATTGCA	TATTATTACT	AACAATAGAA	TTGGATTTAC	2040
	TACAGAACCA	ATTGATGCAC	GTTCAACAAC	TTATTCTACA	GATGTGGCCA	AAGGTTATGA	2100
10	TGTGCCAATA	TTCCATGTCA	ATGCAGATGA	CGTTGAAGCT	ACTATTGAAG	CAATTGATAT	2160
	TGCAATGGAA	TTTAGAAAAG	AGTTTCATAA	AGACGTCGTT	ATTGATTTAG	TAGGTTATCG	2220
15	TCGTTTCGGA	CATAACGAAA	TGGATGAACC	ATCAATTACT	AATCCaGTTT	CTTATCAGAA	2280
	TATTCGCAAA	CATGACTCTG	TTGAATATGT	GTTTGGTAAA	AAGCTTGTTA	ATGAAGGTGT	2340
	CATTTCAGAA	GATGAAATGC	ATTCAATTTAT	AGAACAAGTC	CAAAAGGAAC	TAAGACAAGC	2400
20	TCATGATAAA	ATTAATAAAG	CTGATAAAAT	GGATAATCCA	GATATGGAAA	AGCCTGCAGA	2460
	TCTTGCAATTA	CCGTTACAAG	CAGACGAACA	ATCATTTACT	TTTGATCACT	TGAAAGAAAT	2520
	AAATGATGCA	TTGTTAACAT	ATCCGGATGG	CTTTAACATT	TTGAAAAAGT	TAAACAAAGT	2580
25	TCTTGAGAAG	CGTCATGAGC	CGTTTAATAA	AGAAGATGGT	TTAGTTGATT	GGGCACAAGC	2640
	AGAACAACCTT	GCATTTGCGA	CAATTTTACA	AGATGGTACA	CCGATTCGCT	TAAGTGGTCA	2700
	AGATAGTGAA	CGTGGTACAT	TCAGTCATAG	GCATGCCGTG	TTACATGATG	AGCAAACAGG	2760
30	TGAAACATAT	ACACCTTTAC	ATCATGTTCC	TGATCAAAAA	GCGACATTTG	ATATACACAA	2820
	TTCTCCGCTT	TCAGAAGCAG	CAGTAGTTGG	TTTTGAATAC	GGCTATAATG	TGGAAAACAA	2880
35	AAAAAGCTTC	AATATTTGGG	AAGCACATAA	TGGTGATTTT	GCAAATATGT	CACAAATGAT	2940
	TTTTGACAAC	TTCTTATTCA	GTTCTCGCTC	AAAATGGGGA	GAACGTTTCA	GATTAACATT	3000
	ATTCTTACCT	CATGCATATG	AGGGTCAAGG	GCCTGAACAT	TCATCAGCAA	GATTAGAGCG	3060
40	ATTTTTACAA	TTAGCTGCTG	AAAATAATTG	CACAGTTGTC	AACTTATCTA	GTTCAAGTAA	3120
	TTATTTCCAC	TTATTGCGTG	CACAAGCGGC	TAGTTTATAG	TCTGAACAAA	TGCGACCATT	3180
	GGTTGTTATG	TCACCAAAAA	GCTTACTGAG	AAATAAAACA	GTTGCAAAAC	CAATTGATGA	3240
45	ATTTACTTCT	GGTGGATTTG	AGCCAATTTT	GACAGAATCA	TATCAAGCGG	ATAAGGTTAC	3300
	AAAAGTTATT	TTGGCAACTG	GTAAAATGTT	CATTGATTTA	AAAGAAGCAT	TAGCTAAAAA	3360
50	TCCAGACGAA	TCAGTATTAC	TCGTTGCGAT	TGAAAGATTG	TATCCATTCC	CAGAGGAAGA	3420
	GATTGAAGCA	TTACTAGCAC	AATTGCCAAA	CCTTGAAGAA	GTGTCATGGG	TACAAGAAGA	3480
	ACCTAAAAAT	CAAGGTGCAT	GGTTATATGT	CTATCCATAT	GTTAAAGTGC	TAGTTGCAGA	3540
55							

AGAAATTCAT AAACCTTGTTT AAAATAAAAT TATAGAAAAT GCATTAAAAA ATAAC TAGGG 3660  
 GGAAATAAGT CATGCCAGAG GTTAAAGTTC CAGAATTAGC AGAATCTATT ACAGAAGGTA 3720  
 5 CCATTGCAGA ATGGTTGAAA AACGTAGGGG ATAGCGTAGA AAAAGGTGAA GCTATTCTTG 3780  
 AATTAGAAAC TGATAAAGTT AATGTCGAAG TTGTATCTGA AGAAGCAGGT GTATTATCTG 3840  
 AACAACTTGC AAGTGAAGGC GACACTGTAG AAGTTGGACA TGAATTGCT ATCATCGGCG 3900  
 10 AAGGTAGTGG CAATGCTTCT AAAGAAAATA GTAACGACAA TACTCCACAA CAAAATGAAG 3960  
 AAACAAATAA TAAAAAAGAA GAAACAACAA ATAATTCGGT AGATAAAGCT GAAGTAAATC 4020  
 AAGCAAATGA TGACAATCAG CAACGTATTA ATGCTACGCC TTCTGCGCGT CGATATGCTC 4080  
 15 GTGAAAATGG TGTGAATCTT GCTGAAGTAA GTCCGAAAAC AAATGATGTG GTTCGTAAAG 4140  
 AAGATATTGA TAAGAAACAA CAGGCACCGG CATCAACACA AACAAACAA CAAGCATCTG 4200  
 20 CAAAAGAAGA GAAAAAATAC AATCAATATC CTACAAAACC AGTGATTCTG TAAAAATGT 4260  
 CACGTAGAAA GAAACAGCT GCCAAAAAT TATTAGAGGT ATCTAATAAT ACAGCTATGT 4320  
 TAACAACATT TAACGAATG ACATGACAAA TGTATGGAA TTGCGTAAAC GTAAGAAAGA 4380  
 25 ACAATTTATG AAAGATCATG ATGGTACTAA ATTAGGATTT ATGTCATTCT TACTAAAGC 4440  
 TTCTGTAGCA GCTTTGAAA AGTATCCAGA AGTTAATGCA GAAATCGACG GCGACGACAT 4500  
 GATTACGAAA CAATATTATG ATATTGGTGT AGCTGTTTCT ACAGATGATG GATTATTAGT 4560  
 30 ACCATTTGTA AGAGATTGTG ATAAAAAGAA TTTTGCAGAA ATCGAAGCAG AAATTGCTAA 4620  
 TTTAGCAGTT AAAGCAGAG AGAAAAAACT TGGCTTAGAT GATATGGTTA ATGGTTCATT 4680  
 35 TACGATTACA AATGGCGGTA TTTTGGATC AATGATGAGT ACGCCAATTA TCAATGGTAA 4740  
 TCAAGCTGCA ATCTTAGGCA TGCATTCAAT TATTACAAGA CCAATTGCGA TTGATCAAGA 4800  
 TACAATCGAA AATCGTCCAA TGATGTATAT TGCATTAAGC TATGATCATA GAATTATT 4858

40 (2) INFORMATION FOR SEQ ID NO: 380:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2222 base pairs  
 (B) TYPE: nucleic acid  
 45 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 380:

ATCAGTCACA CGGTAGGCAT ATAAATGAG TCGTTTCTAC AACATTTTAA AACAGTTCAT 60  
 TCAATATTAT TTTTATCTAA TAATATATT GGGAGGATTA TACCTTTATA CACACCATGC 120

55

	ATGCTATTTA GCTAAAGCTA AAAGACCAGA CACTATGCAT ATTTCAACTG GAAATATGTG	240
	GCGATACTTA GTTGCAATTA TTGCCTGTAT GATTTGGTAC CTTAATAAAG CGCATGTAAG	300
5	TATCATCGGT ATAATTATTG GTTTAATGAT TTCATATGTT GTAGTTATCA TACGTCCTTT	360
	ACTAAAGGTG AGCAAATAAA TTAAGAAAGA GGTGAGATTA TGGATCACAA ATCCCCGCTC	420
	GTGAGTTGGA ATTTATTCCG TTTTGaTATC GTTTTCAATT TATCAAGTAT ATTGATGATA	480
10	CTTGTTACGG cGTTTCTTGT TTTTCTACTT GcTATCATTT GTACGCGTAA TTTGAAAAAA	540
	AGACCAACTG GCAAACAAAA TTTCTGTGAA TGGATTTTTG ATTTCTGTGAG GGAATCATT	600
15	GAAGGTAACA TGGCTTGAA AAAAGGTGGT CAATTCCACT TCTTAGCAGT AACGCTGaTT	660
	CTGTACATTT TTATAGCTAA TATGTTAGGT CTTCCGTTTT CTATAGTAAC GAAAGATCAC	720
	ACATTGTGGT GGAAATCACC GACAGCnGAT GCAACAGTGA CTTTAACGTT GTCTACAACG	780
20	ATAATACTGT TAACTCACTT TTATGGAATT AAAATGCGTG GTACGAAACA ATATCTTAAA	840
	GGTTATGTAC AGCCGTTTTG GCCATTGGCA ATTATTAATG TTTTGAAGA GTTCACTTCA	900
	ACATTAACGC TTGGTCTGCG TTTGTACGGT AACATATTTG CAGGTGAGAT ACTATTAACA	960
25	TTACTTGCTG GCTTATTCTT TAACGAACCA GCATGGGGTT GGATTATTAG TATCCCAGGA	1020
	TTAATTGTTT GGCAAGCATT TTCAATATTT GTAGGAACAA TCCAAGCATA TATCTTTATT	1080
	ATGCTTTTCGA TGGTTTATAT GTCACATAAA GTGGCAGATG AACACTAAAA ATTTCAATAA	1140
30	TTATATACAA TCACAGGAGG AAATTAAATT ATGAATTTAA TCGCAGCAGC AATCGCAATT	1200
	GGTTTATCAG CATTAGGAGC AGGTATCGGT AACGGTTTAA TCGTTTCAAG AACAGTTGAA	1260
35	GGTGTAGCAC GTCAACCAGA AGCACGTGGT CAATTAATGG GTATCATGTT CATTGGTGTA	1320
	GGTTTAGTTG AGGCATTACC TATCATCGGT GTAGTAATTG CATTATGAC ATTTGCTGGA	1380
	TAATTAACAG ATAAAAGAGG TCGGGACAAA GCGCATAGGA CATAATTCAT GATGCATATA	1440
40	TAGTAATATC TTTGAACTTT ATTAAATAGT TGAGATATGA ACGCACCATG CCTATCGCAT	1500
	AAATTCAGTA GGTCCTAACC TCGTCGTTTT TTTCTATATA AACTAGCGA TTATTTAAT	1560
	GAAAGGAGTG TCATGAACCC GTGACTGAAA CAGCTAACTT ATTCGTTCTT GGTGCAGCTG	1620
45	GAGGCGTTGA GTGGGGTACT GTGATTGTAC AGGTCCTAAC TTTCATCGTG TTAcTTGCGT	1680
	TACTTAAAAA GTTCGCATGG GGTCCATTGA AAGATGTAAT GGATAAACGT GAAAGAGATA	1740
50	TTAACAGAGA TATCGATGAC GCAGAACAAG CTAAGTTAAA TGCACAGAAA CTTGAAGAAG	1800
	AAAATAAACA AAAACTTAAA GAAACACAAG AAGAAGTTCA AAAGATTTTA GAAGATGCTA	1860
55	AGGTTCAAGC ACGTCAACAG CAAGAACAAA TTATTCaTGA AGCAAACGTA CGTGCAAACG	1920

TTAATAATCa AGTATCTGAA CTATCAGTGT TAATTGCTTC TAAAGTTCTT AGAAAAGAAA 2040  
 TTTCTGAACA AGACCAAAAA GCATTGGTTG ACAAGTATCT AAAAGAGGCA GGCGATAAAT 2100  
 5 AATGGTAAAA GTAGCTAACA AGTATgCTAA AGCATTATTT GACGTGTCAT TAGATACAAA 2160  
 TAATTTAGAG ACTATTAATG AAGAATTAAC AGTTATAAAT GAAGCAGTAA AAGATAAAAT 2220  
 TG 2222

10 (2) INFORMATION FOR SEQ ID NO: 381:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1143 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 381:

TAACAATTCC TaTATTCATG TTTAATACGA AACACTACAT TTACATTGTA ATTCACTATC 60  
 TTTTGAAGTA ATAAAGTGAT TTGTTCAATC GATAGCTCAT TGCTTGTGTC GATTGTAACA 120  
 25 ATTATATGCA AGTTTTTCAGG ATTAACACCT AATCTTTGAA TGATTGTGTTT AATAGTATAA 180  
 TAATCCATCC AATAAAAATT ACTTCCTTTA ATATAAATGT TTTTAGGTTG ATACATTTGA 240  
 CTCCTTTTTG TAGGCTCAAA AGGTATATCA ATCTCGCGCA TACTTGAAGA ACTTTGATTA 300  
 30 GTATCATCAA ATAATTCAAT TATATTTTTA TCAATTTCAA CTAATTGGGA ATGGTTAATT 360  
 GAATGAGATG TTGGACTATA TCTTTTTCTA ATTAATTTAG GAGTGTACC ATACGTTTCT 420  
 TTAAAGAGGT GTATAAACG AGAATAATGA TTAAACCAT GACTACTTGC GATTTCTTCA 480  
 35 ATAGGCTTCT TAGrAGTTAA AATATCAATT AAGCAATGCT CCAGTCTAAT ATGATTTAAA 540  
 TATTGAATAA AATTACTATA AGGTGTCGCT TTAAACATGT CACTTAAAGC TTTGTTTGTA 600  
 40 ATACTAACTT GATTAATGAC ATCTTTCCTA TTTATCTTTT TATGGTGGTT GTTTGTTAAA 660  
 TAATCGTGCA CTTCTTCGGC TACTAAATGA CGACTACCAT CGTATGTATT TAATGACATT 720  
 AATTCAACAC ACATGTAATT AATAATCTTA TCATTAGCAT TATAAGACTG TTGTTTAATT 780  
 45 TTGCTGTAGA TTAAATACTT AATCAAGATT CTAAAAGTGC TAGCAACCTC ATTTGTTAAT 840  
 ATCTTGCCAC GCATCATATA GTTTGCATTC ATAACTTCA AAAATGTCTT TGCACCTACT 900  
 TCAACAACAC AACATACACT ATCATCATTA CCATCAATTT GATACAAATC ATTCAACATA 960  
 50 ATAATGGTGA CATCATTTCT TTTTACATCA TATTGTTGTA GATTGATTGA AACTACACCT 1020  
 GTACCTTGTA GCCAATATAT AATTTTTtAAG TTCGCTTCCT TCACACTACT CATTTTCATA 1080

55

ATT

1143

## (2) INFORMATION FOR SEQ ID NO: 382:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 506 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 382:

15 GGACCACCTT TTTTACCTGT ACGACTAAAG CCAACTGTTT TAAAGGCACC AACAGCATGC 60  
 CACACTTGAA TAATTTCTTG TGATGGTCTA AAGCGAACCG TATAAATTAA TGGATGGAAA 120  
 TCATCAACAA AAATATAATC GGCCTTACCA AGTAAATATG GCAATCTAAA CTTGTCTCTC 180  
 20 CATTTGCGTC TATCCGTAAT ATTCTCCTTA AATACCGTTT TAATATCATA ATCAAAATCT 240  
 ACTTTTTGGC GTAGTAACTC ATCATATACA TACTTGAAAT TCCCTGATAA ATTCGGACGC 300  
 GAATCTGATG TGAATAATAT TGTTTTGCTT CTTTAAATAT GTAGTAACTT TGTAATATTA 360  
 25 AAAATAGCTT TAAATAAGAA ACTTCTACTT TCAAATGAAG CTTTATGACC TTGTTTaTGA 420  
 AGCCAGTGTG cACTTgTCGC AATGaCCCCCT GaTTTCyCyT GagGtAAGGk GaTTTCmATA 480  
 TCAAATACAA ATTGTTAAC GTCAC 506

30

## (2) INFORMATION FOR SEQ ID NO: 383:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 421 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 383:

40 AGCATCTGCA AtGAGTcTCT AATACAACGA TACGTTTTGC ATCTTTAGGT ACTTTTACTG 60  
 TACCATTTTC ATCTTTTACC GAAATAGTAT CTTTAGTTGA TGATTCTTTT TTAATTGAAT 120  
 45 TATCCGTATT ACCACAAGCT GCAACTAAAA GTAAGGCAAC TATTAATCCC AATATACTAA 180  
 AAGTTTTTAG ACCTCTCATC GTTCCACTCC TTAATATGTA TAACTTCATT TATTATTTTA 240  
 TTGATAACAA TTATCATTGT CAAGTAGCGT TCAATCTTTT TTATATTCT AAAATGTATG 300  
 50 ACTATATATT TCCTCTAATA ATTATGACTA CAATTAGCAC ATTTCCCTTAG ACAAATACT 360  
 GATAATGTAT CATTGCTATA TCATCTTTGC ATTAATACAA TTGACACCAT TTAGCATGAC 420

55

## (2) INFORMATION FOR SEQ ID NO: 384:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 862 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 384:

TGCACTTGAT ACTTTAGCAC ATGAAATTGC TAATCGATTA GTTGCAAATG ACAAGAATGA 60  
 AGCAACTTTG GAAATGACTA ATAAATGGC AACGATTCGT TTTACAGAAC CTACGCTGAT 120  
 TGCATTAGCA GGGGGTAATG TCAAAGCTTA CACTGAGCAT ATGACTATAT CTCCATATAA 180  
 ATTGTATTTG TTAGATAAAG GCGATGTTTT AAAGTTTAGA GAAACAAGTT ATACATCGCG 240  
 AGTGTATTTA GCTGTgGGAG GCGGATTGA ATTAGATGCA TGGTTAGGAT CTAACCAAC 300  
 CGACTTTAAT GTAAAAATTG GTGGTTTTAA AGGTAGAACA TTACAAGATG GCGATGAAAT 360  
 AAAGCTTAAG AGAGATTATA CAGCTCGTCA TCATAAGTTA TTTGAAAACC TTGCTCACAC 420  
 GAAACAAACA GATTGGGGTA TTGATGGATA CGCCTTGTC TTTAATTATA TGTCTGATGT 480  
 ATTTTCATGTC GTTAAAAATA AAGGTACGGA AGATTTTAAA GAAGATGCCA TTCAAAGATT 540  
 TGTGAAACAT GATTATAAAG TAACGAGCAA AGCAAATCGC ATGGGGATGA TGCTTGAAGG 600  
 TGAAAAAATC AAAGCTTTTT ATGAAGATAT GCCACCGTAT CAGACTGTCA AAAAAGGAAC 660  
 GATACAAATT AAGCGTGATG GCACACCTAT TATCCTATTA AATGATCATT ATACGCTAGG 720  
 TAGCTACCCG CAAATCGGTA CAATCGCAAG TTATCATTTA ACGAAATTAG CACAAAAACC 780  
 GCAAGGATCA CGTTTGAAAT TTCAATTTAT AGATATTTTA ACGGCTGAAA AGAACCTTGT 840  
 TAAgTATAGT AACTGGTTAA AC 862

## (2) INFORMATION FOR SEQ ID NO: 385:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1027 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 385:

AAATGAGTTG ATGATGGCTC CTATTCCATC TATCTCTAGG TATGACAGAA sATAAACGGT 60  
 CTTCTACCTG TCTsACATTA TCTTTCCAAC GATTAATACC TAAGCGTTTA GAAACACGTT 120

	CAGTTTTACG TCCTACACCT GCTAAACTTT CTAATTCCTT ATGTGTTTGT GGTATTTCTC	240
	CATTAAATTG ATCAATCAAA GATTGACAAA GTTCTTTAAT ATTCTTAGCT TTGTTACGAT	300
5	ACAGACCGAT AGAACGAATA TCATTCATAA GTTCTTCATC ACTGACTGCC AAATAATCTT	360
	CAGGCGTTTT GTATTTTTTA AACAGCTCAG TTGTTACTCT ATTTACTAGA ACGTCTGTAC	420
	ATTGCGCTGA CAATAATACA GCAATAGTTA ATTGGAACGG ATTATCATGT TTTAATTCAC	480
10	ATTCTGCATC CGGAAACATA TTTGCTATAA CATCAATCAT TTCTAATGCT TTTTTCTTAC	540
	TTACCATCAA GGTTCCTCCC GTTTAACCAA TCAAATTTAG GTACCGTTTT AACTGTGTGC	600
15	GTCATTTTCG GTTTATTGAA TTTTCTCTT ATTTTCTAG AATCGTCAAT TGTTTTGACA	660
	TTGTTTTTCT TCCAATTAAG TAAAATACGA TCTATATATT TAAAGCTAAG TTTATTCAAA	720
	CTATTCGCCT CGTCTAATGC CGCTTGTATA ATTGCAGTAT CGTGTTTATC AACATCAATC	780
20	CATTGATTTA ACGTTTCTAT TTCATATGGA GATAACGGCC TTGCAAATGT ATCCTCTAAA	840
	ACTCTAAATA ATTGTTTAAA TTTTCTTTA CTATTTAGCT CTTTCGTTTC CATACTTTGT	900
	TGCTTCAATA TATGACTTAA TTTTTCGAAA AAAGGATCTA GATTCATATA TTCGGKAAAT	960
25	CTACCTTCyT CATCyTTTTG aACTkGtAAT tCTAGCAATT CACGTgTATC AAATTTTGGA	1020
	TACCATT	1027

(2) INFORMATION FOR SEQ ID NO: 386:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1006 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 386:

40	AAGGnTTGGA GGGAATTAAT nGATGGCAAA TCCCAGaTAA AGTACACaAC GCATGGAATG	60
	TGATGCATTT AATATTTCCA GTAGTTAGTA CGACGTTTGC AAGCTTTAAA TCTATGTATG	120
	GGGGCATACC AAAAGATTTT ATAGACTACT TATTTATTGA TGAAGCAGGA CAAGCAATAC	180
45	CTCAAGCAGC TGTGGGAGCA TTATATCGTT CAAAAAAGT TGTAGCTGTA GGTGATCCGA	240
	TTCAAATAGA ACCGGTTGTG ACTTTAGAAA GTCATTTAAT TGATAACATT CGTAAAAATT	300
	ATCATGTTCC GGAATATCTA GTTCTAAAG AAGCTTCTGT GCAGTCTGTT GCAGACAACG	360
50	CCAATCAATA TGGTTTTTGG AAATCTGATG CTAAGTATAG TAATCAAAAA ACCTGGATAG	420
	GCATACCTTT ATGGGTGCAC AGACGATGTT TAAAACCTAT GTTCACGATa GCTAaCCAAa	480



GGTATGACGT TAAAGGAAaC GCAGTTCAAA AACAAATTTGT GAAAGAGCAT GGTGAAAAaG 600  
 TAGTGGGATT ATTAGCTGAT GATTGGATTG AAGCAATTAA GGAAGGTAAA AATGAACCGA 660  
 5 GCTCATTTGT AATATCGCCT TTTTCAGCAG TACAGCAACA GATTAAACGT ATGTTAAAGC 720  
 AACAACTACC GACTAGAATT GATATTGAAC GTACAAAAAT TAATCAATGG GTCGATAAAT 780  
 CCATTGGTAC TGTTCACTACT TTTCAAGGTA AAGAGGCTCA GAAGGTGTAT TTTGTAATAG 840  
 10 GTACTGATAA TACCCAAGAT GGTGCTGTGA ACTGGTCATG CGAAAAACCA AACTTGTTAA 900  
 ACGTTGCAGT GACAAGAGCT AAGAAAGAGT TTaTGTAATT GGCGACATGC AAAGAATACA 960  
 15 GATGAaACCA TTTTATGAGA CGATTTTTAn AGnAAGAAAT GTAAAA 1006

## (2) INFORMATION FOR SEQ ID NO: 387:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 662 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 387:

CGTTTCATCA AGThATCGAT CCTAATTTAC AAAATATTCC GGTTAGACTT GAAGAAGGGC 60  
 GTAAAATAAG AAAAGCCTTT AAACCAACTT CAAAAGATAG CGTTATATTA TCAGCAGATT 120  
 30 ATTCTCAAAT TGAATTGCGT GTATTAGCAC ACATTACACA AGATGAGAGT ATGAAAGAAG 180  
 CATTATCAA CGGCGATGAT ATTCATACAG CAACTGCTAT GAAAGTATTT GGTGTAGAAG 240  
 35 CTGATCAAGT CGATAGTTTA ATGCGTCGTC AAGCAAAAGC AGTTAACTTT GGAATTGTTT 300  
 ATGGGATAAG TGATTATGGT TTAAGTCAAA GTTTAGGTAT TACTCGTAAA AAAGCAAAAG 360  
 CATTCAITGA TGATTATTTA GCTAGTTTCC CAGGTGTAAA ACAATATATG TCTGATATTG 420  
 40 TAAAAGATGC CAAAGCTTTA GGTTACGTGG AAACATTGCT ACATCGTCGA CGCTATATTC 480  
 CTGATATTAC GAGTCGTAAC TTTAATTTAC GCGGCTTTGC TGAACGTACT GCTATGAATA 540  
 CGCCAATACA GGGCAGTGCT GCAGATATCA TTAACTGGC AATGGTTAAA TTTGCTCAAA 600  
 45 AAATGAAAGA GACAACATAT CAAGCTAAAC TATTATTACA AGTACACGAT GAATTAATTT 660  
 TT 662

## (2) INFORMATION FOR SEQ ID NO: 388:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 669 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 388:

5 TCATCCAAAT TTTGGAAATT CCACATTTTA CATATCGTAA TTTTITAGGA AACTAGTGAA 60  
TATAACAAAT CCCTCCTCTC ATTTTAAAA TAGATATATC ACTTCCCCAC TTTTACTTAA 120  
CTAAACTGCA ACGGTTCTTA ATACCAAAAT CCTGCCCTCT ATTTTATCA ATTCAAGCAT 180  
10 ACTTATTGAA AAATGTTAAC GTTTTCTTGA TAATCATTGT AAGCGCATTT ATTTTATAAA 240  
CTAACGTTTG AAATATACTA CAGGAGTGAC ACGTAATGAC TCAAATTACT GAAAAAGAAT 300  
TAAAAAAGAA GTATTTAGAT TTAATATCCC AAAATTTTGA TACTCCAGAA AAACCTTGCAA 360  
15 CTGAAATTAT CAATTTAGAA TCAATTTTAG AATTACCTAA AGGTACGGAA CATTTCTGTA 420  
GCGATTTACA TGGTGAATAT GAAGCTTTCC AACACGTATT ACGCAACGGT TCTGGGAACG 480  
TGCGAGCGAA AATCAATGAT ATTTTCAAAG AGAGACTTTC AACTAAGGAG CTTAATGACT 540  
20 TAACTGCTCT TGTCTACTAT CCAGAAGrCm AATTtAAAAAT TgATTAAAAG TGATTTCCAA 600  
AATTgCGGtC mActTAATGt CyGGtATATC ACaCmATCG aACATTTAAT TGAGTTAATT 660  
AAATATTGT 669

## (2) INFORMATION FOR SEQ ID NO: 389:

## (i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 1249 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 389:

35 CACATGGCTG TTAGAGATAT GAATGGCCAT GCGTTACCTT TAACAAAAGA TGGCAATTTT 60  
TATCAAACGA ATGTAGATGC AAATGGTGTT AATCATGGTG GTAGTGAAAT GGTGCAAAAT 120  
40 AAAACAGGTC ATATGAGTCA ACAArGCCAT ATGAATCAGA ACACACATGA ACCAACAGCC 180  
ACACATGCAA CAAGGTCATA TGCAATCATC AAACCATCAA ATGATGAGTC CAAAAGCAAA 240  
45 TATGCATTCA TCAAATCATC AAATGAACCA AAGTAACAAA AAAGTTTAC CAGCTGCTGG 300  
TGAAAGTATG ACATCAAGTA TTCTTACTGC AAGTATTGCC GCACTACTAT TAGTATCTGG 360  
GTTATTCTTA GCATTTAGAC GACGTTCAAC AAATAAATAA ACATAATACG ATTAATAATA 420  
50 GAAAAATCGT GTGATTATCT GaGGGAGCCT AGGACATAAA TCAATGTCCT AGGctCnCTA 480  
AtGTTATATT GGCAGTAGTT GACTGAATGA AATTGCGCTT GTAACAAGCT TTTCCATTTT 540

55

CGATTGTCTT ATACGTGTCA GTGTAAATTC AGATATTTC TGTGGAATAT ACCACTTATT 660  
 AATCATAATT GGATAAGGTG tTTGTGCGTA CAGTGTTCATA ATAATCAGCC AACAAATGTGT 720  
 5 ATCACCAtCA AACACGTGAC TATGATTTTtk GAAGTGGGGC GCTTTGGTAA TAGACATTTT 780  
 TAAATCTGAT TGATATGCAT TGCTATAAAT CGTTTGCTCA ACGAATGTCT TCATGTCGTC 840  
 TTCGTTTTGT GTATTCACIT TAAATGTGTC AATGACATTT AACGGTATAA AGGTAAAGCA 900  
 10 AAATGCATCA GCTTGCTTAG AATGATTGTC CTTTTTTTGA TAATAGCGTT CCATTGCAAT 960  
 GACGGCAGAA GGATGGTTTG CAAACAAATG ATTTGTATAT TCACTTTCTA AATCAACACG 1020  
 ATAATTAATT GATGACATAG ATACGCGAGC TAGCAATATT TGATCAAGTG GATGCTTAAA 1080  
 15 TTGATCCATA CTGAAGCGT GTTGGGCATT TGTGTGTTGA ATAACAAAGT GTCCCTTCCC 1140  
 TCTGTACTC TCTACGATGC CATCTTCGGC TAACAATTTt ATAGCTTGGC GCAAAGTCAT 1200  
 20 ACGACTGGAC ATCAAAGCGC GCACAAAGTT CTTTTTCAGT AGGTAATGC 1249

## (2) INFORMATION FOR SEQ ID NO: 390:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1788 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 390:

AATGATGATT CATTCAAAGA AATTAACGCT CGGTATATGC TTGGTATTAC TCATTATATT 60  
 GATTGTAGGT TATGTCATTA TGACAAAAAC AAATGGTCGA AACGCCCCAA TTAAGACAC 120  
 35 ATTTAATCAA ACATTAAAT TATATCCAAC CAAAAATCTC GATGATTTTT ACGATAAAGA 180  
 AGGCTTTCGA GATCAAGAAT TAAAAAGGG TGATAAAGGT ACTTGATAG TTAATTCTGA 240  
 AATGGTAATC GAGCCAAAAG GTAAGGATAT GGAAACGAGA GGAATGGTGC TCTATATCAA 300  
 40 TCGCAATACT AGAACCACAA AAGGGTATTA TTTTATAAGT GAAATGACAG ATGACAGTAA 360  
 CGGCAGACCA AAGGATGATG AAAAAAGGTA TCCGGtAAAA ATGGAACATA ATAAAATCAT 420  
 45 ACCAACGAAG CCACTACCGA ATGACAAGTT AAAAAAGAG ATTGAAACT TTAAGTTCTT 480  
 TGTACAATAT GGCAACTTTA AAGATATTAA TGATTATAAA GATGGTGATA TTTATATAA 540  
 TCCTAATGTA CCAAGTTATT CGGCAAAATA TCAATTGAAT AATGATGATT ATAATGTCCA 600  
 50 ACAGTTaAGa AAAAGATATG ATATTCCAAC CAAACAAGCG CCGAACTAT TATTGAAAGG 660  
 CGATGGAGAT TAAAAAGGTT CATCCGTAGT TcTAGAAGTC TTGAATTTAC CTTGTGCGAA 720

55

AGGTATGAGT CAAACTGAAT ATCAAATAAA ACCTGGTAAT ATAACAAGTA ACTCTGAAGA 840  
 AACAAGTTCG ATATCTAAAG TGAGCTGTGA AATATAGGTA GCCATTTCOA AAAATTTAAA 900  
 5 GGTGAATTTG ATAATGTAGC TCAAGGAGAT TGGGTTAAAA AGGCGAAGAA TGAAGTGGAT 960  
 GATATTAGTA AGAAATTAAA AAATATTCAA AGAACGGAAG TTTAATAGCT TATATGATTC 1020  
 TTGGAGCTAA GACAGCATGC GTTCATTCAT GCCATTATTA ATATAAGCAC CGCAACAAAA 1080  
 10 AAGCTTCTAA TGTGATACAG GAACCTCATA TTCCGTATCA TGTTAGAAGC TTTTAATGTC 1140  
 TAAAGAACAT CTACATTTTA TCATATTTTC TGACTTATTA AACTTTTATA TAATTAAATA 1200  
 TTTCTTAATT TTCCAAAATA GTGATAAATT TGTGAAATAC ATCACAAATC CCTTTATTTA 1260  
 15 TTTGGAAATT CATGTAATAT TAGACTTGTA AGAAGTTAAT AAATAGAGAG AGACGAGAGA 1320  
 GTTTATATAA ATACTATATA AACATTGGAG TGATGATTAT GAGAAAAGAG ATTGAAGCGC 1380  
 TTATTTTCTC AGACGTAAAT GCTATGATAT TTACGTGAAC ACTGGTGTAA ATCAAGGATT 1440  
 20 AATTGGTGAC ATCAAAGATG GTTACCTAAC TATTGATTCT ATGCCTTACA TTGATGCTGA 1500  
 GCGTTTGTAT CACTTTGCTA TGGAACGTAA ATCGTTAGTC ACTAACTAGT TCTTATTGCC 1560  
 25 AATGATTACT ACCCCTAGTC GCGGCAATT GAAGTGTGAT TGATGTAATC TGCCCTCGTT 1620  
 GGTGAGCAAT TGAGGGCAGA CCCCTTTAAT TAAGTAAACC CTAACCTCCC ACAAATCTGG 1680  
 AACGATACTA AAAGCCACGT CCTATATTGG ATGTGGCTTT AGTCAKACTT ATATTATTTT 1740  
 30 tAAAACGATT ACCTACAAGA TTTACATATA AAATTCTATC ATGnCTGC 1788

## (2) INFORMATION FOR SEQ ID NO: 391:

(i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 2407 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 391:

GCAAGTTTAG TTAAACAGAT ATTAAAACCA GATCAATATA CAGATGCAAA TTCTCTATTA 60  
 45 GAGATTAGTT TGCAAACAGG TATGTTTATT GCAGGTGGAT TATCAGGAAT ATTATATAAA 120  
 ATAAATGGAT TCACTCTAAT TATAGCGATG ACTATAATGA TGTTTCTAAT CAGCATTTTT 180  
 ATGTTATTTA GATTGCATGT AGATAAACCA ACTCATTCAG AGGAAGAATC AACAAATAGC 240  
 50 TTATTGCAAG AATATTTTATT AGGATGGAAA TTTTAAAAG ACAACATGAT GATATTTTATT 300  
 TTTGGAGTTA TTTCAATCAT ACCAATGGTG TTTACAATGA TCTTTAACAT ATCATTACCA 360

55

	TATGGCATTG GAGGATTATG TGCAGGTTTA ATTTTCAGCTA TTCTTTTCGAA GAAAATCTCA	480
	ACTAAAGTAT TGATATTTTT GTTATATTTT ATATTAGTCA TAAATTCAGC ACTATTTATT	540
5	TGGATAAACT CAGCATTTTA CTTATTCATA GGATCATTTA TACTAGGCTA CTCAATTTCA	600
	TCAATCAGAA TTTATATGAA TACAGCTATA ATGAACACTG TTTCAGATAA ATATGTCGGT	660
	CGCTCATTTA CGATATGGAC GTCAATTTCA TTGTTACTGC AAAGTTTAAT TGCTCCATnT	720
10	TTAGGAAGAT GGATTAATGA AATTAATGAT AAATTCGGTT TCTATATTAT ACTCATTTTA	780
	TCCTTACTCA TATTTGTaCA CTGCTGCTTG TTAACAAAAC AGACAAAATA AAATATGCAC	840
15	ATAAAGAAGA GTGACCGTCA CTCTTCTTTA ACAAGCGACC ATTTATCGaT GGGCTTAGTT	900
	CTCTCTGCAC CCACACTTCA CTACTTCACT TTTTCAAATC ATTTTTTATG GTCTTAAATA	960
	AATCAGTGAG ATTTGTTGCT TCGGTAAAGT CTAGAATTAA TATCATTCTT TTAGAACCTG	1020
20	GATATGGCGA AACTAATGTA TTATCTTGCA ATTTCTGCTG GGCACTTTTA GTCGCCTTGA	1080
	CCAATAATCT ATTATCATAC AAACCACCTA TAACCACGCC ATCATAATAA ATAATATATT	1140
	CTCCCATCAT CTTTCTTGTC TTAACCGCGT TTGAATTCAC ATGATTTAAA AATAAATCAT	1200
25	GTACATCTTT CTTAGTCGCC ATTGTAATCG CTCCTTCAGT TTTATGTTTA ATCACATTCTG	1260
	TATTAATGAT TCATTTCTGTG TTGCTCTTAA TTTTATCTAT AATTATATGC GTAGTTAAAA	1320
	TCAAACCTAT GGAAAAGAAA ATAATGATAA CAATGTTAAG AAATATAGTT ATAAAATTAT	1380
30	AGTTTGGAAG GTATGCGAAT AGCAGAAAAA TAGGTATCGC AAAAAATAAA TCCCACCAAC	1440
	CTAAACTTTT TAAAGAATGC TTTAAACCTT CCATAATATC ACCTTTATAA ATTTGTCTTT	1500
	GTTATAAGAT AACTAAAAAA TCGCTTTACT GTAAAAGTAG CCAAAGAAAA TTCTGAATCA	1560
35	TATTCATAAG TAGTGTATCA TTAATAATGA ACAATTTAAT ACTATAATCC TTGATCTTTG	1620
	TATTGATCAA CTTACCACAA CATTTATTTT AGACTACTCT TAGACTTCCC TTTCAAATGG	1680
40	TTGCATCTAT TGAAATTCCT TTTGTATAAG TTAGGCTTTT GTGGTAATAT CATCATGCAT	1740
	AAAAAATCGA GATACTAATT ATAAAGAGGG TATAAATATA TTATGAAAGA AAATTTTTGG	1800
	AGTGAATTAC CACGTCCATT TTTTATTTTG GCGCCAATGG AAGACGTTAC AGATATCGTC	1860
45	TTTCGACACG TTGTAAGTGA AGCAGCTAGA CCGGATGTGT TTTTCACTGA ATTTACAAAT	1920
	ACTGAAAGCT TTTGCCACCC TGAAGGCATA CATAGTGTGC GCGGACGCTT AACTTTTAGT	1980
	GAAGATGAAC AGCCGATGGT CGCTCATATA TGGGGAGATA AGCCAGAACA GTTCCGTGAA	2040
50	ACGAGTATTC AATTAGCTAA AATGGGCTTT AAAGGCATAG ACTTAAATAT GGGATGTCCT	2100
	GTAGCAAATG TTGCTAAAAA GGGTAAGGGT TCCGGCTTAA TCTTAAGACC TGACGTTGCT	2160

55

GGCTACTATG AAATCGATGA ATGGAAAGAT TGGTTGAAGC ACGTCTTCGA ACAAGACATT 2280  
 GCCAATTTAT CTATTCATCT TCGTACACGT AAAGAAATGA GTAAAGTAGA TGCACATTGG 2340  
 5 GAATTAATCG AAGCTATTAA AAATTTACGT GACGAAATTG CACCAAATAC ATTGTTAACA 2400  
 ATTAACG 2407

## (2) INFORMATION FOR SEQ ID NO: 392:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2424 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 392:

20 ATGATGnATG GcncCGCCAA GAAGTTGAAC CAGTCTATTG GCTAGAATTG CTCAAAAAGC 60  
 GAGAcATGTG GTATTCaTAT GTTaGTAgtCt ACGCAAAGAC CATCTGTCAA TGTAATTACA 120  
 GGTTTAATTA AAGCCAACAT ACCAACAAGA ATTGCATTTA TGGTATCATC aAGTGTAGAT 180  
 25 TCGAGAACGA TATTAGACAG TGGTGGAGCA GAACGCTTGT TAGGATATGG CGATATGTTA 240  
 TATCTTGGTA GCGGTATGAA TAAACCGATT AGAGTTCAAG GTACATTTGT TTCTGATGAC 300  
 GAAATTGATG ATGTTGTTGA TTTTATCAAA CAACAAAGAG AACCGGACTA TCTATTTGAA 360  
 30 GAAAAAGAAT TGTTGAAAAA AACACAAACA CAATCACAAG ATGaATTATT TGATGaTGTT 420  
 TGTGCATTTA TGGTTAATGA AGGACATATT TCAACATCAT TAATCCAAAG ACATTTCCAA 480  
 ATTGGCTATA ATAGAGCAGC AAGAATTATC GATCAATTAG AGCAACTCGG TTATGTTTCG 540  
 35 AGTGCTAATG GTTCAAAACC AAGGGATGTT TATGTTACGG AAGCAGATTT AAATAAAGAA 600  
 TAATTATGAG TAAGGAGTTT TATATAATGA CACACTATCA TTTTGTCCGA ATTAAAGGTT 660  
 CTGGCATGAG TTCATTAGCA CAAATCATGC ATGATTTAGG ACATGAAGTT CAAGGATCGG 720  
 40 ATATTGAGAA CTACGTATTT ACAGAAGTTG CTCTTAGAAA TAAGGGGATA AAAATATTAC 780  
 CATTTGATGC TAATAACATA AAAGAAGATA TGGTAGTtAT ACAAGGTAAT GCATTGCGGA 840  
 45 GTAGCCaTGA AGAAAtAGTA CGTGACATC AATTGaAATT AGATGTTGTA AGTTATAATG 900  
 ATTTTTTAGG ACAGATTATT GATCAATATA CTTCAGTAGC TGTAAGTGGT GCACATGGTA 960  
 AAACCTTCTAC AACAGGTTTA TTATCACATG TTATGAATGG TGATAAAAAG ACTTCATTTT 1020  
 50 TAATTGGTGA TGGCACAGGT ATGGGATTGC CTGAAAGTGA TTATTTGCT TTTGAGGCAT 1080  
 GTGAATATAG ACGTCACTTT TTAAGTTATA AACCTGATTA CGCAATTATG ACAAATATTG 1140

55

TGGCACATAA TGTTAAAAA GGTATTATTG CTTGGGGTGA TGATGAACAT CTACGTAAAA 1260  
 TTGAAGCAGA TGTTCCAATT TATTATTATG GATTTAAAGA TTCGGATGAC ATTTATGCTC 1320  
 5 AAAATATTCA AATTACGGAT AAAGGTACTG CTTTGTGATGT GTATGTGGAT GGTGAGTTTT 1380  
 ATGATCACTT CCTGTCTCCA CAATATGGTG ACCATACAGT TTTAAATGCA TTAGCTGTAA 1440  
 TTGCGATTAG TTATTTAGAG AAGCTAGATG TTACAAATAT TAAAGAAGCA TTAGAAACGT 1500  
 10 TTGGTGGTGT TAAACGTCGT TTCAATGAAA CTACAATTGC AAATCAAGTT ATTGTAGATG 1560  
 ATTATGCACA CCATCCAAGA GAAATTAGTG CTACAATTGA AACAGCACGA AAGAAATATC 1620  
 CACATAAAGA AGTTGTTGCA GTATTTCAAC CACACACTTT CTCTAGAACA CAGGCATTTT 1680  
 15 TAAATGAATT TGCAGAAAGT TTAAGTAAAG CAGATCGTGT ATTCTTATGT GAAATTTTTG 1740  
 GATCAATTAG AGAAAATACT GGCGCATTAA CGATACAAGA TTTAATTGAT AAAATTGAAG 1800  
 20 GTGCATCGTT AATTAATGAA GATTCTATTA ATGTATTAGA ACAATTTGAT AATGCTGTTA 1860  
 TTTTATTTAT GGGTGCAGGT GATATTCAAA AATTACAAAA TGCATATTTA GATAAATTAG 1920  
 GCATGAAAAA TCGGTTTTAA TATGTTTATA ATAGAGTAGT ATGGGTATTT ATTATTAATG 1980  
 25 ACATTATTAC ATGTTAATTA GGAGGCGTTT TTAATGGATT GGATTTTACC AATTGCTGGA 2040  
 ATTATCGCTG CGATTGCATT CTTAATTTTA TGTATCGGTA TCGTAGCTGT ATTAAATTCT 2100  
 GTTAAGaAAA ACTTAGATTA TGTGCAAAA AACTTGACG GTGTAGAAGG TCAAGTTCAA 2160  
 30 GGTATTACTC GTGAAACAAC AGATTTACTT CATAAAGTAA ACCGTTTAAC TGAGGATATC 2220  
 CAAGGTAAAG TAGATCGTTT AACTCAGTT GTAGATGCTG TTAAAGGTAT CGGTGACTCA 2280  
 GTACAAACGT TAAACAGCTC TGATAGTCGT GTAACAAATT CAATTACACA TAATATTCTT 2340  
 35 CAAATGAAG ATAAAATCTC ACAAGTTGTT CAATGGTCAA ATGTTGCAAT GGAAATTGCA 2400  
 GACAAATGGC AAAATAGACA CTAC 2424

(2) INFORMATION FOR SEQ ID NO: 393:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 738 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 393:

50 CnATAATATT CnTCAAnCCT TTAAATAGAG GAATAGACTG CCGACAGAGT CCGAGACTTT 60  
 GTGGGTAGTT TTTTtagTTT TGATAACGGA AGTTAGAGGC TCTCTGTCAA ATTGGGCAGA 120

AGTGTAAGT TTTCAACATA ATACTATTAG TTCGGTCATG TATCGGACTG ATGGAAAAGC 240  
 GTTTCACCTT TAATGACTCA TTAAGAACGG CCTGAAAATG TTTGGCGTAT TAAGTGCAAT 300  
 5 GATAGTTTTG ACATTAGT TCTAATTGGT CATTACTGCC GAGCAAATCT AGTAGAGTAA 360  
 TCATGTAAAT CTTTAATGTG CCATTGATT CACTAGCGGT GTTAATAACT ACGGAAATTG 420  
 CATTTCGAC TGAAATTTTT GAAAAATATC AACGTACGCT ACAAATAAAA TTTTAAACTG 480  
 10 TTATAAATGT GTCTCAATTT CATATGTTCA TCGACGATAT GAAGCGTATT ATGGTAAAAT 540  
 GAAGAAATAA TAACTTGTT AATAAATAAA ACATCAGCAT TTGACTAAAG CACTTTATTA 600  
 15 TTGTGTAGAT AATAGTTTT TAACGAAATA AAAATGGCGA CTGGTTTAA TAAATCAGCT 660  
 AATGAATCAC TACACCTATA AGTATGAATA TAGTGATTAG AATGCTTGT ATAGTTGGAT 720  
 TTGCAAAAT TGATGTTA 738

20 (2) INFORMATION FOR SEQ ID NO: 394:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1270 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 394:

30 AAAAGTTGTA ATTAAAAGTG GGATTTTACT TAAGnnAGAA GGAACTATT TATATGACTA 60  
 ATAAAAGAGn AGATGTCCGC AATATAGCAA TTATTGCTCA CGTTGACCAT GGTAAAACAA 120  
 CTTAGTAGA TGAGTTGTTA AAACAATCTG GTATATTCAG AGAAAATGAA CATGTCGATG 180  
 35 AACGTGCAAT GGA CTCTAAC GATATCGAAA GAGAGCGTGG AATTACGATT CTAGCCAAAA 240  
 ATACGGCTGT TGATTATAAA GGTACACGTA TTAATATTTT GGATACACCA GGACATGCAG 300  
 40 ACTTTGGTGG AGAAGTAGAA CGTATTATGA AAATGGTTGA TGGGGTTGTC TTAGTAGTAG 360  
 ATGCGTATGA AGGTACAATG CCTCAAACAC GTTTTGTACT TAAAAAGCG CTAGAACAAA 420  
 ACCTGAAACC TGTTGTTGTT GTTAATAAAA TTGATAAACC ATCAGCACGT CCAGAGGGTG 480  
 45 TTGTAGATGA AGTTTTAGAT TTATTATTG AATTAGAAGC AAACGnTGAA CAATTAGAAT 540  
 TCCCTGTTGT TTATGCTTCA GCAGTAAATG GTACAGCTAG CTTAGATCCT GAAAAGCAAG 600  
 ATGATAATTT ACAATCATTA TATGAAACAA TTATTGATTc ATGTACCAGC TCCAATTGAT 660  
 50 AACAGTGATG AGCCCATAC AATTTCCAAG TAGCATTGTT GGA CTACAAT GATTATGTTG 720  
 GACGTATTGG TATTGGTCGT GTATTGAGAG GTAAATGCG TGTCGGAGAT AATGTATCAC 780

55



GATTAAAACG TTTAGAAATT GAAGAAGCAC AAGCTGGAGA TTTAATTGCT GTTTCAGGTA 900  
 TGGAAGACAT TAATGTTGGT GAAaCTGTAA CACCACATGA CCATCAAGAA GCATTGCCAG 960  
 5 TTCTACGTAT TGATGAGCCT ACTCTTGAAA TGACATTTAA AGTTAACAAT TCTCCATTTG 1020  
 CTGGCCGTGA AGGTGACTTT GTAACAGCAC GTCAAATTCA AGAACGTTTA AATCAACAAT 1080  
 TAGAAACAGA TGTATCTTTG AAAGTTTCTA ACACAGATTC TCCAGATACA TGGGTAGTTG 1140  
 10 CTGGTCGCGG TGAATTGCAT TTATCAATCC TTATTGAAAA TATGCGTCGT GAAGGTTATG 1200  
 AATTACAAGT TTCAAAACCA CAAGTAATTA TTAAAGAAAT AGATGGTGTA ATGTGTGAAC 1260  
 CATTGAACG 1270

## (2) INFORMATION FOR SEQ ID NO: 395:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1365 base pairs  
 (B) TYPE: nucleic acid  
 20 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 395:

AAGACCAGGA GAACAmGTAA AACAATATAC AGTTGaATC GTCGTAAAT TAATGGAATT 60  
 TGATATAAAA TGCTCGTGAT TGCTTGTAATA TACGCAACTG CTGTnGCTTT AGAATATTTA 120  
 30 CAAAAGACCT TATCAATCCC CAGTGATTGG CGTAATTGAc CAGGTGCTAG AaCAGCAATA 180  
 ATGACTACTA GAAATCAAAA TGTATTAGTA CTAGGAACGG AAGGCACAAT TAAATCTGAA 240  
 GCATATCGTA CGCATATTAA ACGTATCAAT CCACATGTAG AGGTACATGg CGTTGCCTGT 300  
 35 CCAGGTTTTG TGCCACTTGT AGAACAAATG AGATATAGTG ATCCAACAAT TACAAGCATT 360  
 GTCATTCATC AAACACTGAA ACGTTGGCGT AATAGTGAGT CTGATACTGT CATTTTAGGA 420  
 TGTACCCACT ATCCATTGCT CTATAAACCT ATCTATGATT ATTTTGGTGG TAAAAAGACA 480  
 40 GTGATTTTCGT CTGGATTAGA AACGGCTCGT GAAGTTAGTG CATTGCTAAC ATTTAGTAAT 540  
 GAACATGCAA GTTATACTGA ACATCCAGAT CATCGATTTT TTGCAACAGG TGATCCTACT 600  
 CACATTACTA ACATTATCAA AGAGTGGTTA AATTTATCTG TCAATGTGGA ACGTATATCA 660  
 45 GTGAATGACT AGGAGGATTT TTAATGAAAG AGATTGTTAT TGCATCGAAT AATCAAGGGA 720  
 AAATAAATGA CTTTAAAGTA ATATTTCCAG ATTACCACGT AATAGGTATT TCAGAACTAA 780  
 50 TACCAGATTT TGATGTGGAA GAAACAGGAT CAACATTTGA AGAAAATGCT ATATTAAAAT 840  
 CAGAAGCTGC TGCAAAAGCA TTGAATAAAA CGGTCATAGC TGATGACAGT GGACTIONAAG 900

55

GCGATGAAGC AAATATTGAA AAATTATTAA ATAAGCTTGG TAATACAACT GATCGTCGTG 1020  
 CGCAATTtGT TTGTGTCATA AGTATGAGTG GCCCTGATAT GGAAACAAAA GTATTTAAAG 1080  
 5 GTACTGTTTC AGGTGAAATT GCAGATGGAA AATATGGCGA AAATGGTTTC GGATATGATC 1140  
 CGATATTTTA TGTACCGAAA TTAGATAAAA CCATGGCTCA ACTTTCAAAA GAACAAAAAG 1200  
 GGCAAATTAG TCATAGACGA AATGCGATTA ATTTACTTCA AGCTTTTCTT GAAGGTGATA 1260  
 10 AAAATGTCTA AATGGATTAT TGTGAGTGAT AACCATACTG AATCAGGCGT TTTATATCAA 1320  
 ATTTATGAAA TGCACCCAGA TGCAGATGTA TATTTACATT TAGGA 1365

(2) INFORMATION FOR SEQ ID NO: 396:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1383 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 396:

25 AATTCCTGGT GCAATAATAA ATAGGATGAA AAAGATTCGG AAAATATGAT AACTCGTAAT 60  
 CATAGCAACA TCGGCACCAG TAGCTAATGC AACTAAACT ATCTGATTAA CCCCTCCTGG 120  
 TGCTGCACCA AGaACAATT CATTAATAGG ATTATTATCA AAGAAATGTA TGATATAAAC 180  
 30 CATGATTAGC GCACCAATTA TCAACATAAT ATTTTGAATT GTAATTGCGA TTGCTAGTCT 240  
 ACCTTTTAAA TCTGACAATA AATGCGCAAT TTGAACTCCA ATTCTAATCA TATATATTAG 300  
 TTGTGCCATG TTCAACAACC AATGATCTAG TGTAATGTT AAACCTGTAG AAAAATTCCA 360  
 35 AACAATTAAT ACAATGAGTG GTGCTAATAA TTGAAATGTT GGAAACTTTA TTTTAGACAT 420  
 AATTAGATAA ACTATAAAGA TAGCTATCGC TAAAATAACT ATTTGCCCTA TGTTTAATAC 480  
 TTGTGATAAA GGCAAGACTT TTGTTAACTT TCCATTGCGA TGCATGTTAC CATCATGAAA 540  
 40 AAAATATGAA ATGAACGGTA CTAAAACAAC AACAAATATA ATTCGTGATG TTTGCGTTAA 600  
 GCTAACAACT AACAAATTAG CACGTTTGTC TTGTTGAGCC ATGACCAGCA TTTGTGTTAG 660  
 TGCTCCTGGT ATAACACTTA AAATAGCTGT TTCTGTATTA ATACGTGCAA TTTTTTAAAA 720  
 45 AACAAATGCC ATTACTATTG CAATTAATAA TATCGAAATA GATACAACAA TAATCGAAAG 780  
 CCAATTGTTT TTAATATCCA TAACGACATT TTTGTAAC GTTGATCCGA TTTGCACACC 840  
 50 TAATAGTACA ATACCTAATT CACTAAGTAA GAATGGCCAT TTAATATCAA GTTGAAAAAC 900  
 TTTTACACAA ATGATTGATG CGATAATAGG ACCAAACATA AATGGAAGTA ATACGTGCGA 960

55

TATCATTGCC ATGTTTTCCA CTTCTTTCAA TAAAAAATAA AATGACTAAA TTGCTGCTTG 1080  
 AGCTTCACGT TTGTTAAGAT AACAAATATCC GCTAGCAGTT tTGACTACAA AGCATATATG 1140  
 5 GaCTTTCaCT ATCAAGTCGC CGCCCATGCC TTATATACAT TTAAAnGAG CCTGAACAAA 1200  
 GTTCAGGCTC TCAATTTGTC CGTATATTTA TTTTACAATA CGACTTAAAG CCGTATCAAA 1260  
 TGCTTGAATC GTTTTTCAAT ATCTTCTTTC GTGTGTGCCG TAGATAAGAA TGTACCTTCA 1320  
 10 AATTGAGATG GnGGnAAAAA CACACCTCTT TGnCATTCTC GGTACATTTC TGCAATAATT 1380  
 TCC 1383

## (2) INFORMATION FOR SEQ ID NO: 397:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 415 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 397:

25 TCCACTAAnA TGATTACAAT TGCATTAGTT TGGTGGAGTG CATTTACAAT CTTAACGGGT 60  
 ATGATTAAAG ACCACGGTTT AATTaTTTAG TGAGATTCTT ATTTGGTGTA GGTGAGGCGC 120  
 CAATGTACCC TTCTAATGCT GTGTTTAATT CATTTTGGTT CTCTAAAAAT GAAAAAGTA 180  
 30 GAGCATCAAG TGCATTATTA GCAGGATCAT ATTTCCGACC TGTATTAGCA CCAATAGTTA 240  
 CAATTGCTAT TGTTAACGCA TTTAAGTGGC AAGCACTATT TTACATTTTT GGTGCAGTAG 300  
 GTATTTTAAT GGCKGTATTa TGGGCGATTA TTGCCAAAGA CTTACCTGaG CrACATAGwa 360  
 35 TGGTTAATGA AGCGGAGAAA CGTTTCATTA TGGAAAATCG TGATATCGTA GCTAC 415

## (2) INFORMATION FOR SEQ ID NO: 398:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1141 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 398:

45 TTTTAGaTaA aTyCAATTnT CyATaCTaAA TgATTnTCTT ATTaCGTCAA TTcGCCTTTT 60  
 50 aTTTTATCGT AATCTTTCCa CTGCAAAGCT AAAGCTTCTC CTATTCTAAG ACCAGAATAA 120  
 AATAACAGTC TAGTTAGCTG ACGAGAAGTA TCATTTGTGA TTTGTCTAC TTTTTCATCA 180

55

EP 0 786 519 A2

AATGTGGGGT CGTATAAGAG CTTGTAATGC TTTTGGCGT AATTGATAAC TGCTTTAAAA 300  
 CCTGCCACA CAGATCGTGC ATAGTCAACA GAAAGACCTG CATCGTTTAA CAAATAATTC 360  
 5 CTGAAAGCAG TACATTGCGT AGTAGTGATT TTGCCAATAG GGATATTTCC GAACCTTTCT 420  
 TTTATGTGAG TATTATATTC TGTAGTTCGC TTTTCTATTG AGCGTGCAGA AAGATTTTCA 480  
 TTTTTTAAAC GATCAAAAAA TATATATTCA AAGGGTTGAT TGTCCGAGTA TCCATATTTA 540  
 10 ACATTTTGTA TAAATTCGCT TTCAGCTAGT TTGGCATCTT TCTTACGTTT AAACCCACGC 600  
 TTCATTTTTC GTTGTGTTATT ACCGTATACA TCTTTATATC TAATGGAAAA ATACCATTTA 660  
 CCTGTATTAT CATCCTTATA TACTGGCATT TTGCTTCTCC CTCCTCAAAA TTGGCAAAAA 720  
 15 ATAATAAGGG TAGGCGGGCT ACCCGAAATT TAGTACTAGG TACTAAATGT GATATAATAA 780  
 AATAAAAAGT AGGTGATGTT ATGACATTTA AAAACAATCA TAATTTCAAT GAATTAGTTT 840  
 TAACGAATGA AGACATTAGA ATTTTAAAAA ATGTCTTAGA AGATGCAGTC AGTGTTTATG 900  
 20 ATGAATATTC GGTATGTAAT GAAGAATCCG ATTTTGCTTA CTGTTTATTA AGAGACTTAT 960  
 ATACATTAGA CAGCTTAGCT ATTCGTCAA ATAATGTTTG AATTATCGAA TTGTACTCTT 1020  
 CGATTTTAAT ACCATGCATA ATAGAGTTTC TGTGTTCAAT AGCAGCTTTG ACTGAATGtK 1080  
 25 TTAAATGTTT TTCTATTAAA TCGTTGTTTt CCAtTTCGtK TAAAAATGtT CyTATATTC 1140  
 T 1141

30 (2) INFORMATION FOR SEQ ID NO: 399:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 706 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 399:

40 ACTGTGGTAG GTTTTTTATT TTGAAGTATT AATCATAACA GACTAATAAT CATGAGGTAA 60  
 CTAATAACAC ATATTTAACT TGTATTCTTA AACTGGTATA ATAAATTTAT GTTGAAATGA 120  
 ATATTGTATG ACAGGGTATT CACTTTTATT AAAAGGTAAA ATTAAATAAA GGTTTTATAG 180  
 45 AACGTATTTA AATATATGAG GAGTAAACAA ATGGCTGATA GAACGAATAA AGAAATTAAA 240  
 ACAGGACGCT TTATTGCAAC TGCATCAATC GTATTCTCAA TATTATTGAT TATTCATTAC 300  
 TTTGTTTCGT TGGATAATGC GACTGCCAAA GCATTACTTA ATTTAACGAA TCAAAACACT 360  
 50 TCAGATAAAG CGATTGATTA CATTTTAAAC AGCTTTAGAT TCACTGGTAT TATGTATATT 420

55

ATGTTTGCAG TTTATGTATC AAATAGTTTG TTTACGTTGA TTAATTTATC AATCACAATT 540  
 CAAGCAATAA AAGCTGCACA CGGTGCGTAC TTAACATTGC CAATTTTAAT TGTATTATA 600  
 5 GGTTCGGTTG CATTAGCGAT TTATATGCTT GTTGTTTCTA TCAAACGTAA AAGTACATTT 660  
 AATCGCTAGA AAATTGATTT TAACAATAAA AATATGAAAA AAAAnn 706

(2) INFORMATION FOR SEQ ID NO: 400:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1187 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 400:

ACACAATCTG AAGATTCACG TTGTGGTGCT GGACATGATC GAAAAATTAG AGCTGAACAA 60  
 ATGAnAGAAA TCAGTGATTT TGTAAAAAG AAAAATATCC CTAAAGATGA AACGGTATAT 120  
 ATAGGTGGCG ACCTTAATGT CAATAAAGGC ACTCCAGAGT TCAAAGATAT GCTTAaAAC 180  
 25 TTGAATGTAA ATGATGTTCT ATATGCAGGT CATAATAGCA CATGGGACCC TCAATCAAAT 240  
 TCAATTGCGA AATATAATTA CCCTAATGGT AAACCAGAAC ATTTAGACTA TATATTTACA 300  
 GATAAAGATC ATAAACAACC AAAACAATTA GTCAATGAAG TTGTGACTGA AAAACCTAAG 360  
 30 CCATGGGATG TATATGCGTT CCCATATTay aCGTTTACAA TGATTTTTCA GATCATTACC 420  
 CAATCAAAGC CTATAGTAAA TAGTGCTCAA CTAACATAA ACTTGCTtCG TTCTAAAAGG 480  
 ACGAAGCGAG TTATATTGTT AAAATTTGAA TTGACTTACA TTTTAATAAA ATCATCTTAA 540  
 35 CAACTTTAAT TTTTCaTTAA TACAaGTCTT TACTCTACAC TCAAACnAGA TTCATACACT 600  
 GCACGTCATA ATAAATCTAT CTATTCAAAT ATAAATAAAA GTTACCTACT ACATTCTATG 660  
 40 TAGCAGGCAA CTTTTATTAC TTATTTCTTT TCATTATCAT TAAGTACTTT TACAAACTTC 720  
 ACATTATGTG TCTTCCAATC AACTTCATAT AATGCTGATA ATTTTTCTTC TTTTTTATCT 780  
 ACATGGTTTT CACCAGACCA ATAGCCCCAG AAACCATGGC GATTCCAATC TATTTTAAAC 840  
 45 TCATCCATTG ATCTTTTATA ATGAACAACA AATTGTGATT TACCTTTGTC TTTTTTATCA 900  
 TGTGACATAA CAGCTAAAAA TTCTGGATTA AACCCCTTCAG ACACAGTTAC AGGCATTTTG 960  
 TCTTTAGGTG TGAAATTATC TTTCGCCCAT AAATTTCCAT TTCGTGTTAA AGAAAAAATT 1020  
 50 TCACTTTTAG TTCTATTATC ACTATCATTG GTTAATTGTC TCGTATGGTC ATGTCCCATA 1080  
 TTATTTATCA AATGTGCTTC TACTTTCCAA CCTACACCTT TATGTGACGT AGATTGATCA 1140

## (2) INFORMATION FOR SEQ ID NO: 401:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 847 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 401:

CCAGAATTAT TTTTCAAAA AGGACAATT T AACAATGTCG ATAACGTTAT CATAAGCAAA 60  
 CCGATGAAAG GGACAATGCC TAGAGGTAAA ACGGAaGCTG AAGATCAACA GTATTATAAA 120  
 ACATTGCAAA CTTCTTCGAA AGATCGTGCA GAAAATGTCA TGATTGTTGA TTTACTAAGA 180  
 AACGATATAG GGAGAATATC ACAGAGTGGC TCAATTAAGG TGTATAAACT ATTTTTTATT 240  
 GAGGCATATA AAAGTGTATT TCAAATGACT TCGATGGTAA GTGGAAC TTT AAAAAATAAT 300  
 ACAGACTTAA CTCAAATTTT AACATCGTTA TTTCTTGTG GTTCGATTAC AGGTGCACCG 360  
 AAAGTGAATA CAATGAAATA TATTAAACAA TTAGAAAGTT CACCTCGTGG TATATACTGC 420  
 GGACAATTGG ACTATTACTT CCAACTGAAG ATGATAAAAT GATTTTTTAAT ATTCCGATTC 480  
 GCACTATTGA GTATAAATAT GGACAAGCGA TTTATGGAGT CGGAGCAGGT ATTACAATTG 540  
 ATTCTAAGCC AAAAGATGAA GTGAATGAAT TTTACGCAAA AACCAAGATT TTGGAGATGT 600  
 TATAATGCAA TTATTTGAAA CAATGAAAAT TGATAATGGA CATATCCCTA GACTTACTTA 660  
 TCATACTAAT CGCATAAAAT GTTCTTCTGa GCGATTAAAC TTTAAATTTG ATGAACATGC 720  
 ATGGCGAAAT GAATTAAACG ATGTAACAAC AAAGTATCAC AGTGGTCAAT ATAGACTTAA 780  
 AATCGTATTA AATGCTGAAA GCAAATTTGA AACGATAGTG TCACCTTTAC CTGAGAAAAG 840  
 TAGTTTT 847

## (2) INFORMATION FOR SEQ ID NO: 402:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 740 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 402:

TGAAGATGAA GCAGAAGCAG AAGACAATCT GCTACGAGTA CAATCGAAGA AAGAAGAAGT 60  
 GTATCGTCGA TTAATTGCTT CTAAGTTAAC AAGCGTTCCT GAAAGtTTAT CATCATGAAA 120

GTTAAACAGT TAAAAGATAA AGTGTCTAAA ATTGTGATTC AAATGAATAC ATTTGAAGAT 240  
 GAAGCAAATG ATGTTCTTGT TAATGCTGTT TATGCAGAGA AATTAATTCA ATATGGAAAT 300  
 5 AGATATCGTA AGGACTATAG CAATGTTGAT AAGAGCTTAA ATGAAGCTGA ACGATTATTT 360  
 AAAAATAATC GCTATAAGCg TCGGATTGAA ATTGCAGAGC AAGCTCTTGA AAGTGTGAG 420  
 CCAGGTGTTA CTAAACATAT TGAAGAAGAA GTTATTAAGC AATAGAACT AGTATGTAGT 480  
 10 TATACTTAAA TAATATGAGC ACTCTGTCAA ATTGGACTGA TGAGTTTAAT AATTGAAGTT 540  
 AGCCAACGAT ACGTTGTCTA GCTTCTTTTT TATATGGATA AATGaAAGGG ACAATAAATA 600  
 TAAATAGCAA TTGTTTAAAG ATAAACGTAA TCAAATGTGT TGTTTAAAT AATATAAGTA 660  
 15 GTGAAAAAAG CATAATCACA CAGCTGTTTA AATAGAGTGA AATAGTCTAA TTCTTATTTA 720  
 ATAAGTAGAA ATAAGATTAT 740

20 (2) INFORMATION FOR SEQ ID NO: 403:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 630 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 403:

ATGCCAATAA TTAAACCATG TAATAAATAT ACGTATAGCG TACGACTACC AATATAAGTA 60  
 TATAATTTTT TCTTTGTTGA CATTAAATTT AGAAACGCAG TCATTGCGAT TAATATAATT 120  
 35 CCATATAATA TAAGTCGTTT AAAAGGACTG AATATACTCT GTCCTTCATT TTCAAGTGAA 180  
 GTATATGGTG AACTTCCCAA TAACCAATCT GCATTGATAG GATGAATCAC GTAAACGATA 240  
 AAAACAAAA TAAAGGTAAT GATAGATACT GGTATTAGTT TTTATTTTT AAAAATAGCC 300  
 40 GTATGTTTTT TGGTGAAAAT GTAACCTAGA TAAAATATTG GGAAAAATAC GATTGTCCTT 360  
 GAAATGCTTA AGTAGCTATC GATGTTATCT GAAAAACCTG CTCCAATAGA TATAATAATT 420  
 GAAACTGATA GCACTTTATA TGGATTAAAT CTTCTAACTA TTAATAAAT GACATGAAAG 480  
 45 AAAAATAGCG TGATCAAAAA CCATAACGCA AATACTGGGT TAAAAGGATC AAGTTGTAAT 540  
 TCGTCACTTT TACCTGTTAA GAAATAATAA ATTGAAAAGA ATGCAAAAAA TATCATATAA 600  
 GGTACTATCA AACGTTTTGA AATTTTTTCT 630

50 (2) INFORMATION FOR SEQ ID NO: 404:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6254 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 404:

	ATTTAAACGT TCAGTTTCTA AAAGTGAAC AATCCAAGAA GTACGTAAAC GTGAATTTTA	60
10	CGAAAAACCA AGCGTAAAC GTAAAAAGAA ATCAGAAGCT GCACGTAAAC GTAAATTCAA	120
	ATAATTAATA CCTCTGTTGA CTCCCTCAAC ACGAATATTA ATTATATAAA ACAAACATCA	180
	CAAGTTAGTG TCTGACACTA ATATGTGATG TTTTTTTGTT GTCAATTTTT AATTAAAAAA	240
15	AGTTATATAG TTTATAAATA ATCAAATTGA TATTCTATAG GTTCTTATAA CTATAAAGTA	300
	TATTCAATTT CATGTATAAT TAATGTGAGG GCGAGGTGAA ATTGTGAGTT ATAATAATTT	360
	TTTACAAATG ACAACTATCT TGGAATCAAC GGCTGGAGAT ACTTGGGTTG AACAAGTTAG	420
20	CAATATAATT GTTCAACCTA TTTTACGTT AATATTAACC TGTTTGACAT TCTTAGGATT	480
	TGTATATCAA CTTTACTCTA AAAAAATCAA TGCAGCTGGT ATTATCGCCA CATTATCATT	540
25	ACTTATTTTA TTTTGGGAT TTCTAATCCA AGGAAATGTC AATATGCATT CTATCTTAAT	600
	ATTCTCAATT GCGGTTATAT TAGTTGTAAT TGAATTATTT GTAGTTGGTG CAGTAATTGG	660
	TATTATTGGC ATGATACTGA TAACTATAAG CATAACAACG CTCGGTGATA ATTTGCTATT	720
30	TATGCTTGCG AATGTTATCG TTGCCTTGAT TTTAACGATT GTAGAATGGG TGATATTAGT	780
	GAAGATTTTC AACAGAAAGA TTCCGTTTTT GGATAAAGTT ATCTTAAAAG ATTCAACTAA	840
	TTCTGAGTCA GGTTACAATT CTCATGATAA CCGCTCGCAC CTCGTAGGAA AGACTGCTCA	900
35	AACAGTTACA GATCTTCGAC CTGCAGGGAT TATTTTTTGT GAAAATGAAC GTATTGATGC	960
	TGTTTCAGAT GGCAACTTTA TTTTGCAGCA TAAAACGGTA AAAATCCTTG AAGTTGAAGG	1020
	AACAAGAGTA GTTGTGAGGG AAGTAGATTA ATTAAAAGGA GCGATACCAT GTTTAGTTTA	1080
40	AGTTTTATCG TAATAGCAGT TATTATAGTA GTTGCATTAC TTATTTTATT CTCATTTGTA	1140
	CCCATTTGGT TATGGATTTT AGCGTTAGCA GCTGGCGTTC ATGTTGGTAT AGGTACATTG	1200
45	GTTGGTATGC GTTTACGTCG TGTATCTCCA AGAAAAGTTA TAGCGCCATT AATTAAGCG	1260
	CACAAAGCAG GACTAGCATT AACAACAAAC CAATTAGAAT CGCATTATCT AGCAGGAGGA	1320
	AATGTTGACA GAGTTGTTGA CGCTAATATT GCTGCACAAC GTgcTGACAT TGATCTTCCT	1380
50	TTCGAACGTG CTGCTGCAAT TGaCCTTGCA GGACGTGACG TATTAGAAGC GGTTCAAATG	1440
	TCTGTTAATC CTAAAGTCAT TGAAACACCA TTTATCGCAG GTGTAGCAAT GaACGGTATT	1500
	GAAGTGAAAG CCAAAGCTCG TATCACAGTT AGAGCTAATA TTGCTCGACT TGTTGGTGGT	1560

55



	AGTAAGCATC ATACAGAAGT ACTTGAAAAC CCAGATAATA TTTCTAAAAC AGTTTAAAGC	1680
	AAAGGTTTAG ATTCAGGTAC TGCATTTGAA ATTTTATCAA TTGATATTGC TGACGTTGAT	1740
5	ATTAGTAAAA ATATTGGTGC AGACTTACaA ACTGAACAAG CATTAGCAGA CAAAAATATT	1800
	GCACAAGCAA AAGCTGAAGA ACGTAGAGCT ATGGCTGTAG CAACTGAGCA AGAAATGAAA	1860
10	GCGCGTGAC AAGAAATGCA TGCTAAAGTA GTTGAAGCCG AATCTGAAGT ACCATTAGCT	1920
	ATGGCTGAAG CATTACGTTT AGGTAATATC AGTGTAAAG ATTATTATAA TTTGAAAAAT	1980
	ATCGAAGCTG ATACAGGCAT GAGAAATGCA ATTAATAAAC GAACTGATCA AAGTGATGAT	2040
15	GAGTCACCTG AACATTAAGT CGAGAGGTGA TTAAATGAGT GTCGGTATTC TAATTTTGT	2100
	CATATCAGTG ATCATTTCTA TCATTACTAC TATGCGCGAA AATAGTCATA AAGATAGACA	2160
	AAATCAAAAG CCACCTCAA AAACATCTAC CGATAATGAA CCAAAAAAAG GTGGCTTTTT	2220
20	TGAAGAAATT GAGCGAACGT TTAAAGAAAT AAGTGAAGAA TTAAATGAAG AAGAAAAGAA	2280
	ATCATCGAAA CGAAATATG ATGATACGTT ACCACCTTTA TTCGATGAAC TTCCAAAGGA	2340
	AGAGCCTAAA TCGAAACCTG TTGTAGAACC TATGGCACCT AAAAAACAAC AAGAAACAAA	2400
25	ACCGATGACA GAGAAACCAA TCACAGTGCC TAAAGCAGAA CCGGTGGAGC AGAAACATAG	2460
	ACCTTCTAGA CAAGATAATT CTGACGAAAT TAGACGTCAA TTAGAAAAAT CACTTAGAGA	2520
	TGATATTAAA ACGATTCGTA CTGACATTGA TAGAGAAAAA GAAAAGCAA TTGCTAAAAT	2580
30	GGAAAAACGT GCTAGAGATA TTATTGAGGA TAAATACTTA TCTGAACGTA CAAAACGTTT	2640
	GAAATTAAAG CAGCTGCTTA ATTCTCAAAA TGTCGAAAAA GATTTGACTA AATCAGCGTT	2700
35	CCAATTTGAT AAAGATGAAG TAATCAATGG TATGATATGG TCAGAAATTT TAGCTAAACC	2760
	AAAACAATTA TAAATTTTT TGAAAAAAG CACTATCGTA ATGGTAGTTG CTTGTTTTTT	2820
	TACGTTAAGG AAAATTAAAA AACAAAGAGA ATTTTTCGAG AAATATTAGT TATTTAAATT	2880
40	ACAGCAAAAA ATTGATTAGT CTAAAATTGA ATCTGCTTTT ATGACAAGGT GAAAAGTATA	2940
	AATGATTATT TTAAATTAAA GAAAAATGAG TAAGTCAATG CAAAGATGTT TAAATCAATC	3000
	AATTGCATGA TATAATTAAG TAGATATTAA AGCATCATAG AATGAATATA AATGATATAT	3060
45	GAAAAGGAGC GCGTGATGC CTGGAATTAT ACAAATAGAC GATATGAACC AATCTCAAGC	3120
	TTTAATTGGA AATAATGATG AACATTTAAA AGCAATTGAA GAGAGTTTCG ATGTTGTCAT	3180
	CCATGCAAGA GGACAAGaAG TTGCCGTTAA AGGTACAAAA ATAGAAAACG TAGAAAAAGC	3240
50	GGAATCAGTA TTAATCAATT TGCTGAAGGT TATTGATTTA GGTAATAATA TTACAATTAA	3300
	AGATGTTGAA GCAGCTATTA AAATGGCGCA TAATAACACA ATTCAACATC TGTTAGATTT	3360
55		

	GCAACGTATA TATGTTAATG CCATGAAAAA TAATGATTTA GTATTTGGTA TAGGTCCTGC	3480
	TGGTACAGGT AAGACATTCT TAGCTGTAGT TTATGCAGCA AAGCAACTCC GTAAAGGTGC	3540
5	TGTTAAACGT ATTGTATTAA CAAGACCTGC TGTGAAGCA GGAGAGTCAC TTGGATTTTT	3600
	ACCAGGAGAT TTGAAAGAAA AGGTAGATCC ATATTTAAGA CCTTTATATG ATGGTCTATA	3660
10	TACTGTTCTT GGGCGTGAAC AAACAGAGCG ATTTATTGAA AGAgGCATTA TCGAAATAGC	3720
	GCCACTTGCA TATATGCGCG GACGAACATT AGAAGATGCA TTTGTAATTC TTGATGAGGC	3780
	GCAGAATACG ACACATGCGC AAATGAAAAT GTTTTAAACA AGACTAGGTT TTGGCTCAAA	3840
15	AATGGTAGTT ACTGGTGACC AAACCTCAAAT CGATTTACCT AAAGGTGTTA AAAGTGGAAT	3900
	TAAGGAAGCG GTCAGTAGGT TACACAACGT TAAAGGTATA AGTATATTGA AATTAGATCA	3960
	GAGCGATGTA GTCAGACATC CATTGGTAAG TAAGATCATT GAACATTATG AAGGAGAGAA	4020
20	TTAAATGTTT ACGATAGATT TTAGCGATCA CACAGGCTTA GTTAAAGATG CTTGGTATAA	4080
	ACAAATTGAA GATTTATTAG AATTGCTAA AAAAGAAGAG CATATAGAAG ACGATGCTGA	4140
	GCTTTCTGTT ACATTTGTAG ATAAACAAGA AATACAAGAA ATTAATCGAA CATATAGAGA	4200
25	TAArGwTAAr GTTmCaGATG tAaTCyCaTT tGCTTTAGrA GrAGATGAGC CmGaGATkGA	4260
	TtTTAGTGGT CTTGATATAC CACGTGTTTT AGGGGATATA ATTATCTGtA CGGATGTAGC	4320
	GCAAGrACAA GCAAACAATT ACGGACATTC TTTTGAACGA GAATTAGGAT TTTTAGCATT	4380
30	ACATGGATTT TTGCATCTAT TAGGTTATGA TCATATGACT GAAGCGGATG AAAAGGAAAT	4440
	GTTTGGTCGA CAAGATACAA TATTAAACGC ATATGGATTA ACACGAGACT AATTATGAAA	4500
35	AGGTTTAAAT ATGCACCTGA TGGGCTGAAA ATCTTAATTC AAAAAGACTA TAAATTTCTT	4560
	TTACATGTGT TTGCAATGAT TGTGCTATT GTCTTTGGTC TCGTACTAAA TATTAATCGG	4620
	ATTGAGTGA TATTTATACT CATTGCTATT GCATTAGTTC TCACTGTTGA AGCTTTAAAC	4680
40	ACTGCTATTG AATATGTTGT CGATTTAGTG ACCGTTGAAT ATCATGATTT AGCTAAATAC	4740
	GCTAAAGATA TTGCGGCTTT TAGTGACTT ATAGTTTCAA TATTAGCATT TATTATAGGT	4800
	TTAATAGTAT TTTTACCACA TTTTATAGCG TTATTTTAGG GAGGCATATA TGAGTTATCA	4860
45	ACCTCATTAT TTTCAAGAAG TTAGAAAAGC ACAACAAGAA TCATATTCGC CATACAGTCA	4920
	ATTTAAAGTA GGGGCTTATT TAAAAmCGAA AGACgGTAGA ACTTTTTATG GTACCAATGT	4980
	AGAAAATGCT TCTTATCCAT TATCGATATG TGCTGAACGA GCTAGTTTGG TATCGGCAAT	5040
50	TTCTCAAGGA TACAGACCAG GTGATTTTGA ATCAAtAACT GTAACCGTAG ATGCAGATAA	5100
	ACCGTCATCA CCTGTGGTG CATGTCGTCA AGTTTGAAG GAATTATGTG ATGATGATAT	5160
55		

	ACCATTGGA TTTTCAGGAA AGGATTTAGA ATAAATGACA GAACATAAAT CAGGATTTGT	5280
	TTCAATTATA GGTAGACCAA ATGTAGGAAA GTCAACATTT GTTAATAGAG TGATCGGCCA	5340
5	TAAAATAGCA ATCATGTCCG ATAAAGCTCA AACAACTAGA AATAAAATTC AAGGTGTTAT	5400
	GACAAGAGAT GACGCGCAA TTATATTCAT TGATACGCCA GGTATTCATA AACCTAAACA	5460
	CAAATTAGGT GACTATATGA TGAAAGTCGC TAAAAATACA TTATCTGAGA TAGATGCAAT	5520
10	CATGTTTATG GTTAATGCCA ATGAGGAmAT TGGACGAGGC GATGAATATA TTATAGAAAT	5580
	GTTGAAAAAT GTTAAGACAC CAGTATTTTT AGTATTAAAT AAAATAGATT TAGTGCATCC	5640
	AGATGAATTA ATGCCAAAGA TTGAAGAATA TCAAAGTTAT ATGGACTTTA CAGAGATTGT	5700
15	ACCTATTTCA GCATTAGAAG GGCTAAATGT CGATCATTTT ATTGATGTTT TAAAGACGTA	5760
	TTTACCCGAA GnACCTAAAT ATTATCCAGA TGATCAAATT TCAGACCATC CTGAACAATT	5820
20	TGTAGTGGGT GAAATCATTC GTGAAAAAT CCTTCATCTT ACAAGTGAAG AAATCCCTCA	5880
	TGCGATTGGT GTTAATGTGG ACCGTATGGT TAAAGAAAGC GAAGATCGTG TTCATATCGA	5940
	AGCAACTATA TATGTTGAAA GASGTTTCGCA AAAAGGAATT GTCATTGGAA AAGGCGGTAA	6000
25	AAAGTTAAAA GAAGTAGGaa AAcGTGCGAG ACGTGATATA GaAATGctTC TAGGCTCTAA	6060
	AGTTTACTTA GAATTATGGG TCAAAGTTCA AAGAGACTGG CGAAACAAAG TTAACTTTAT	6120
	TCGCCAAATT GGTATGTTG AAGACCAAGA TTAATCTTAA AAGTGGTGAA GATAATTGTT	6180
30	AATGCGCCAA AAAGGGATTA TCATCAAAGC AGTTGATTAT GGTGAATCTG ATAAAATTAT	6240
	CACGATTTTA AATG	6254

## (2) INFORMATION FOR SEQ ID NO: 405:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3710 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 405:

45	GTGTGTTCTAA ATGTTTCTTG nATGAAACGA GTCAATGTAA ACTGATATTG CTGTATTTGT	60
	TGCAGCAATT CATATTGGTC TGGTGTTGCA ATAACAGCAG CTTGAGTTGG AGTCGCAGCT	120
	CTGATGTCTG CAGCAAATC ACTTAATGTA AAGTCTGTTT CATGACCAAC TGCTGATATA	180
50	ATCGGTGTCT TACAATTATA TATTGCACGG ACGACAGCTT CTTCGTTGAA ATTCCATAAA	240
	TCTTCTATGG ATCCACCGCC TCGACCTACA ATAATGGTAT CTACACCTAA ACTATCTGCA	300

	TGTATTTGTT CAGCTAATGG AAAACGACTA TTTATCGTTG AATGGATATC TCGAATTGCG	420
	GCACCTGTAC TCGCTGTTAA AACTGCAATT TTTTATAGGAA ACTTAGGTAT TGATTTCTTA	480
5	TTGCTTTTAT CAAAACAACC TTCTTCAGTT AATTTTTTCT TTAATGCTTC TAATTTTTGA	540
	TATAAGTTCC CTATACCATC TAATTGCATT TTATTTACAT AAATTTGATA GTTCCACGA	600
10	CGTTCAAAAA CAGAAACACG TGCTTCTAAT AAGACTTCAT CTCCTTCTTT AGGTTCGAAG	660
	TTTAATTTAG AAGCACTACC TTTGAACATC ATGGCACTTA TAACGCTTTC TTTATCTTTC	720
	ACATTAAAGT ATAAATGACC ACTTGaATGc TTTTGTAAAT TTGAAAGCTC ACCTTTAATC	780
15	AATACAGATT GGAGATGTGG ATCTTGATCA AATTTATATT TAATATATTT CGTTAAAGCT	840
	GAAACACTTA AATAATCTGA CATATAACAT CACTCAATTT TATTTTTTTA TATTACTCAA	900
	TACACCATT ATAAATTTAT AATGATCATC ATCACTGAAT TGTTTTGTTA ATTCAACTGC	960
20	TTCAATTCATT ACGACTTTAG CAGGTGTATC ACTGTGTAAT ATTCATATG TTGCCATTCT	1020
	TAAAATAATA CGATCCGTTT TTAATAAACG TGCAATAGTC CAATCTTTTA AATAAGGACT	1080
	AATTGTCTCG TCTAATACAG GTTCGTGATC TTTAACGCCA GAAACTAGCC AATGAATAAA	1140
25	TTGGAAGTCT AAATCTGGAT TATCGTCTTT AATAAAGCTT ATCGCTTCAT TTATCGTTAA	1200
	ATCACTGTCC TTCATTTCTA ATTGAAATAA AGTTTGAAAA GCTTGCACTC GGGATTCTTT	1260
30	ACGACTCATT TTTAACTCCT TCAAACGTTT GTATTTTTCT TTATTTAATT ACTGAATTAG	1320
	GTATGACATT ACTTTTCAAT AACGATTTGT GTAATGTGAA TATTAATTTG CTTAGGTTCT	1380
	ATCGCTGTCA TATTAGAAAT TGAATTAAAA ATTGACGTTT GAATTTTGTT TGCAGTTTTT	1440
35	GAAATATTAA CACCATGTTT TAATGCACAA TATACATCTA TATATATGCC ATCTTCTTTA	1500
	CTCTCGATTT TTAAATCACG GCTTAAATTT TTACGACTAA CTTTTTCTAA ATTTGTTTCT	1560
	TTTAATTCAG CAAAATGGCC AGTGATGCCT TCGACTTCCG AAGTAGCTAT ACTTGCAATA	1620
40	ACAGATAGCA CTTCTGGCGC TATTTCTACT TTACCTAATT TTGAATTTGA ATAATCAGTT	1680
	ACTTTGACCA TGGATTGACC TCCTATTAAC CTTTCATCATT CATAATGCTA TTTTGCTCTA	1740
45	AAAAGTTTGT ATTAAATTTA CCGCTTCTAA ATATATCGTT ATTCAATAAT TTAATATGGA	1800
	ATGGAATAGT TGTATCAATA CCAAGAACCA CAAATTCACT TAGTGACGA ATGCCAGCCA	1860
	TAATCGCTTC ATCTCGTGTC GGTTTCATGTA TGATTAATTT CGCTACCATC GAATCATAAT	1920
50	ATGGCGGTAT CGTATAATTA GTATAACATG CTGACTCTAT TCGAACACCA TATCCACCTG	1980
	GTGCAAGATA TTGCTCmATT TTACCTGGTG ATGGCATAAA GTTCTTGTA GGATTTTCAG	2040
55	CATTAATTCT AAATTCaATT GCGTGTCTCG TTAATTTAAT ATCTTCTTGT TTATACGGTA	2100

CAGTTACAGG ATGTTCTACT TGAATACGTG TATTCATTTT CATAAAATAA AATTTATTAT 2220  
 CATTTAATC ATATATAAAC TCAATTGTTC CCGCATTTTC ATAATTTACA GCTTTCGCTG 2280  
 5 CACGAAGTGC GGCATTTCCC ATTTACAGAC GTGTTTCATC ATCTAAAATT GGGGAAGGTG 2340  
 CTTCTCCAC TAATTTCTGC ATACGTCTTT GAATGTGACA ATCAGGTTCT CCTAAATGAA 2400  
 TTACATTACC ATAGCTGTCC CCAACAATTT GGATTTCAAT ATGGCGGAAG TTTTCGATGA 2460  
 10 ATTTCTCCAT ATAAAGTCCA CCATTACCAA ATGCAGTTTG AGCTTCTTGT TCTGTCATTC 2520  
 GGAAGCCAGT TTCAAGTTCT TTTTCATCAC GAGCAACACG GATACCTTTT CCGCCACCGC 2580  
 15 CAGCAGTAGC TTAAATGATG ACCGGATAGC CAATTTTTTT GCGGATTTTC TTAGCTTCTG 2640  
 AGACGTCTTT CATTAAACCG TCACTACCAG GAACAACTGG AACATTGGCT TTGATCATTT 2700  
 CTGCCTTAGC AACATCTTTG ATACCCATTT TTTGGATAGA TTGATAACTT GGTCCAATGA 2760  
 20 ACTTCAATTG GcATgctTCG CATAATTCTG CAAATCAGC ATTTTCAGCT AAAAAGCCAT 2820  
 AACCCGGATG AACGCCATCA CAACCTGTAG AAGTTGCAAT AGATAAGATG TTCGGAATAT 2880  
 TTAAATATGA ATCTTTAGAC AAAGTGGGAC CTACGCAATA TGCTTCATCA GCAATTTGAG 2940  
 25 TATGTAGCGC ATCTTTATCC CCTTCAGAAT AGATTGCAAC AGTTTGGATG CCTAAATCAC 3000  
 GACAAGCGCG AATAATCCTA ACTGCGATTT CACCGCGGTT TGCAATTAAA ACCTTTTTTCA 3060  
 TTATTTTACC TTAAATAACG GTTGGCCATA CTCTACCATT TGTCCGTCTT CTACTAAGAT 3120  
 30 TTCAACAATT TCACCTGAAA TTTCTGCTTG AATTTCAATTA AATAGTTTCA TTGCCTCTAA 3180  
 AATACACACT GTTGTTCAT TTGAAACAGT GTCCCAACT TGCACATATG CTTCTTCGTC 3240  
 TGGAGATGGC GATTTGTAAA ATGTACCTAC CATAGGTGCA TTAATTGTTT TGTGATTATC 3300  
 35 TGAAGTTGGC TTTGGAGCTT CAGTTTTATT GCTATCAGTT GATTGTGCTT GAGGCATAGG 3360  
 CATTGCCGCA GCTTCAACTG GCATTTGTGA GATTTGTGGC GTGATAATCT CAGTTTCTTT 3420  
 40 TTCTTTCTTA AGCGTCACTT TGCCTTAGT ATCTTCAATA TTGATTCCG TTAAAGTTGA 3480  
 TTTATCCAGA ATTTCAATTA ATTCTTTGAT TTCTTTAAAG TTCATTATTA CTGACTCCTT 3540  
 CAGTTTGTTC TCATCTACCC GTCTATTTTA CTGAGACAA CTCTTCAATT CAAGCATGTT 3600  
 45 CATATTGCTG GCGACATTAT AAGTCTATCC CAAAGTTATA ATAAAACCAC ATTTTAAATT 3660  
 AAAAACACTT GTGTATTTAT TACTTAACAT TGAATCATCT TAACTCTTGA 3710

50 (2) INFORMATION FOR SEQ ID NO: 406:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1705 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 406:

5	GCTGATGTTT GTTGCCTTTn TCCACCAGAC AATTCAGAGG GATATTTATC ACTAATATCC	60
	AATATATTTA ATGCTTCTGC TACTTTTTCA TAACGATTTA ACATATGTTT TTTATCTAAC	120
	TTCTGTACCG TTAGTGGTAA CATTATGTTT TCTTTAACAG TCAATGTATG CAGTAAATTA	180
10	TACTCTTGAA AAATAAAACC AATATCATGC TTGCGTATAT CAGATWATTC CTTGTTTGAA	240
	AGCTTTTCTA ATTTTTTTCC TTTAATGTA ATAGAACCCT GTGAAATATA ATCAATTGAA	300
	CTTAAACAT TTAATAATGT CGTTTTCCCA GATCCAGAGG GACCCATAAT AGCAATAAAC	360
15	TCGCCTTCTT CAATAGACAT ATTGATATCT CGCAACACTT CTTGTGCCAT TTTTTAGTT	420
	CCATATATTT TTGTTAATTG TTTACTTCT AAAATTGCCA CTTTAACACT CCTATAATTT	480
	ATCTTAACCT CATTTCTTTT AGGCTTTGGC ACTTGTATCT TCAATTTAAC ATATGACTAA	540
20	CATCTATCTT ATTATAACGT TGAAGCTGCA TTGATGTATC AATTCTAAGT AACAAAACGC	600
	ATGTTTAAAA TGACAAATTT GTCACCTCCG ACATGCGTTC AACAAATTTCA TTTTGTAATG	660
25	GGAAAATCAA TCTGACAGTT GTCCCTTAC CAACAGTCGA CGTGACTTGC AGGTGAATAC	720
	CTAATTGATC CTTTACACTA TTTACTAAAT ATAGACCCAT ACCTGAAGAC GTCGTTTCAT	780
	TTCTGTTAGC CGTTGACGTA AATCCTCGTT CAAATATTCG CGGCATATCT TTTTACTAA	840
30	TACCTCTGCC ATAGTCTTTA ATATATAACG AAACATGTTG ATCATTTAAT TCTGTCCCAA	900
	TTTCAATATT AAAATTCTCA CTATATTTCA ATGCGTTTGA CAAAATTTGT CTAATAATCA	960
	TACGACACCA TTTTATATCT GTATAAACAT AATCATCCAC TTTAAAGTCA ACATCAAAC	1020
35	CAATACCTTT AACCTGACTA ATATGTCTTG TTAATTGTAT TTCATCAATG ACCATGCGTT	1080
	TAAGTGACAC GTAATCAAAA TACATATCTT TACGTTGAGA TTCTAATCTA GTAATATACA	1140
40	GCTGTGTATC TAGCATCGAG TTTATACGAG ACCATTCATA TAGTAATGCT TGTWTTCTTT	1200
	CTTGATTTTT TTCTTGATCA ATTAATAATT TCATAGCTGT CACAGGtGTT TTTATGTCGT	1260
	GCACAAATTC TGTAATGGTT TGTTTCATGCA TGTTCAATTG CAACTGTTGC TCAACAACCT	1320
45	TTTCTTTGTG CGCTGAGATT TGACGATATA AATAATCAAC TGTATGACGT TGAAATGGCG	1380
	TTTCCGCTAA ATCTTTATGT TTAATTTCTT CTATTTCTTT ATCTTTGTCA AAATGCTTAT	1440
	ATAATTTTAC TTCTTTAAAA TATGTCAATA AAAGAAAAAT CATTGTTAAA CTTAAATTCA	1500
50	AAGAAACAAT ATAAAATAAA CTGTCTATTG GAAAATCATA ATCGATTAGA CTATGTCCTA	1560
	ACATAAGGAA GTTTAAAAAC AATATCCAAA AtATCCAGTT CATGCGAGAT TTCAAAAAAT	1620

55

GCACTATCCA TACTAATTTTC AGATA

1705

(2) INFORMATION FOR SEQ ID NO: 407:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1722 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 407:

15 TCAATATATC TTAAAATTCA ATGATTAAAT CnATTATCAC TAGACATnAA ATACATAAAAT 60  
 CCTATTCCCC ATTTTCATTT nTTAATTCAT AAATGAATCA ATAACCACCT AATAACAAAT 120  
 CATATTATAC ACCTTTGTTC TCTATTTTTC TAAGGTTTAA AAAATATTTT TAGGTAAACC 180  
 20 TAAAAATAGA TGTAATAAAA ACGCCTCCTC AGATAATTTAT ATATCTATGA AGACGTTTAA 240  
 ATACATTATA GATGGTCTGG TTCTGGGTGA ACGTATACTG AGGAAALACC TTTTTTGTGC 300  
 AAATGATGTT CGACATTGTC ACAAATTTGA TGCCTTCTA CTAAGGAAAG GTTAGCATCT 360  
 25 ACAACAATTG TGACATCAAT AAACACACTA CTTCCATGGT AACGCCCTTT AATACTTTTA 420  
 ACTTCTTGTA CTTTCATCAAC TTCTAAAATA TCATTGCGAT ACGCTTCTAA TTCAGTTTCA 480  
 TTGAAACCAT CACTCAACAT AAAAATTGCT TCTTTAAAAA TACCAAAACC AGTATAAACG 540  
 30 ATTAGTAAGC CTAGTAATGT TGCTAAAATA ATATCGACAA TTGGGAAACC GATTTCGCTA 600  
 AAAATTAATC CTATCGCTGT TCCAATGCTG ACTAACTAT CCGATAAATT ATCTTTGGCA 660  
 35 GCCGAATTTA AAGAACTACT TTTCGTTCTT TTCGCTAGTC TTTGATTGAC TGCAAATACA 720  
 ATCAACATTA CAAGACCACT GATTAAGCTG ACGATAATTG TTATTGCGTT AGGTACAACG 780  
 TCATCTTCTT TGAACAAACG AGGTGCATTT TGAATAACTA CTTGGATACC TACAAACATA 840  
 40 ATGACAAATG ACACCAATAA TGAAGAAATA TTTTCAGACT TCAAATGGCC ATAAGGATGA 900  
 TTTCGATCGG CAGGTTTAAT TGAAATTTTC AATCCAATAA TAACAGCTAA AGAAACGATA 960  
 ATATCTGTCA TATTGTTTAA TGCATCGGCT CTTACAGCTG CAGAGTTAAA GACAAAACCC 1020  
 45 GTGACATACT TAACAATAGA TAAGATTATA TATACAATTA AACTCAAATA AGCACC CGCT 1080  
 TGCGCCAATT TAAGATTTTC ATTATGAGAC ATGCGTTGAA CCACCTTGAA TTAGTATAGT 1140  
 50 AACAATATTA TGAATGATTC ATTTTAATTT TACAACGTTT TTAATTTTTA TAAATTTTTA 1200  
 TAAAAATAAA CTAATTTATT CATTGCAAC CCCTAAAAAT AATTTTTTAGC CTTTCTGCGA 1260  
 ATTTTATGAG CTAGAAAGGC GCCCAACTCT CCCTGTTTGT TAACTTTCGC CTCGAAAGTT 1320

55

CGAAwTTAT GAGCTAGAAA GGCTTATGCA GTTGACGTTT TACGTCCAAC TCGGTTCCCTC 1440  
 CGTCTTCTTC AAATTTATTT GTnAGAAAGG CACCCAACCTC TCCCTGTTTG TTAACCTTCG 1500  
 5 CCTCGAAAGT TTCTATGTTA GAACCCTATG CATGAGTTGC GAAnTATCTA ATGTCGTGAA 1560  
 CTAATTATAT AGAAGAAAAA GTGCATCAAT GACAAATTAA ATGAGATTTC TACTCTACCA 1620  
 AACTCTCTTC GAAAGACAAT TTTCTCCTCT ATTTATTAGC AACTATTGCA TTTCTCCATA 1680  
 10 TAGTACTTCC TTACTIONAAAA TACGCTGAAT GTCTGAATTA AA 1722

## (2) INFORMATION FOR SEQ ID NO: 408:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 5521 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 408:

GCGAGACCCC CTGAGGGAGC AGTGCCAGT CGAAGACCCG AGGCTGAGAC GGCACCCTAG 60  
 25 GAAAGCGAcC ATTyCAATAC GAaTTgTGat AAATAGAGAA CAGCAGTAAG ATATTTTCTA 120  
 AITGAAAATT ATCTTACTGC TGTTTTTTTTA GGGATTTATG TCCCAGCCTG TTTTTTGTGA 180  
 TTTTAAATAA TTTGAATATG GAAAATGTAT TATTCTCTCA TTTGTATAGA TTGTATTTAA 240  
 30 TAAGTTAATG TAATCCTTGA GCTCACGATT AATAAAATTC TATAACCTTA ATTATTTTCT 300  
 CGATACAAAG GGTATTAAAC TTTAATATAA GTATAATGAT GTGCCTCATC TTCAAGACGC 360  
 35 ATTGTTGTAA CACACTTATA ATCTATAAAT GGCGCGAACA TGGTATCTTT AATTTCaTTT 420  
 ATGCGATTCT CATTTACTTG ATTAGATTTG TGTGTCGAAA GTACAAGTTG ATCaAAAATG 480  
 TTATCTAGTA CATCACGAAC GATATACCAC ATATGTCTTT CTAAGTTTGA ATCATTTGAT 540  
 40 GCTTTAGAGA TTGTAAGAAT TAATTCGCCT AAATGGTTTT GAACGGTAGA ATAAAAGGCT 600  
 TTGTTAAACA CAGACGTTTT TGAATCAGTA AGAATTCTTG ATTTTTCATG GAAATGAGAT 660  
 GTRACTGTATA CCATTTCAAT TAGTTGTGCT TTATCAATTC TTAAACCTTC AAAGTCTCTA 720  
 45 ATATACATCG TGTCCAATAG ACCATCTTTC CGAAATGTTG CAATAGCATT TTGCAAGTGA 780  
 GCCTCTAATG CAATGCCATA TTTAGTAACT AGTGGGATTA CGAGACCAAG CAATGCTTTA 840  
 CTATAAGTTT CAACCCACGA TTTCGCCGAT GATTCAAAAT CAGACAATGA TGCAGCTGAT 900  
 50 TGATAACGTT TAATCAATGT CACGATAGGT GATTCAATGT TAAATGGGTA GGTGCAACT 960  
 AAGCTTGAAG GAATCAATGG TGTGACTTCT TGTGGAATCA TTTGGTATAT ATTTTTTCTA 1020

55



EP 0 786 519 A2

	AAAAATGAAT ACCAGCAACT TCATCAATAA TTGTTGATGC ATAGGACTTA AATATGACAT	1140
	CTTTCTCCAA AATATCATTT AAAATACGTG TCATTAGTGG ACCATTGTGC GTCGTTTGT	1200
5	CTGATAATGT ACGAATCTCA CCTGTAATAT GAACGTTTGT CGACAATTG ATGTGTGGCG	1260
	ACATAGCTGG GTATTTAGGA ACTAATGTTT TGAAAGATAA ACCAGCATAA TAATCCaACG	1320
	TATGTTTTGC TTCAATGATT AATTCTTTAT CTAATTCTGC TTGATAATCA GAATGTAATA	1380
10	CGTCATCTAA TTGCCATGGA TGAACAATCA TAATGTGATA ATCATTAAAG TTAAACTTTG	1440
	GCGTAAATTC ATTTTCTAAT TGTTTAATTA AGTCCGGAAA TAGTTGATGA ACAGTTGTAT	1500
15	CATAATCTTT AGACAGTGAC ATAGTACGGC TTAATTTACT GTGAATCAGT ACTATTTTCA	1560
	ACTTAATAGG TTGATTGAAT TCTGAAGAAT ATAGGAATGT TTGTAATGCA TTTAAACCTT	1620
	TACGTAATTT AGCCCCAGGA TGTAGCGGAT GACCTTCAAT AACGGCTTGC TCTGAACGCA	1680
20	AGTAACTATC TTCGCTATTT TCGATAATAT TAAATAAAGG TGCAGAATCA TGTTGCATTG	1740
	ACAGTGCTTG ATAGCTAATT GCAAATGTCA TATtAGTtGC ACTGTTTATT AAATCTTGCT	1800
	GAAATTGATC ACTAGCAGCA TTTTTTAAAT CTGGTGCTTC AATTAAAATA CACTCAAGAA	1860
25	TTTCATTTGG ATGGTGACT CGTGAATCG TATTTGTAAT GTCATCTTTA ATGTAGAAAG	1920
	GGCCTTCAAC ATCAATTCGA TCAAAGGCGT GTTCTCCAGT GATAGGAGCA TATAATGTTT	1980
	GCTTAGCTTG TGGGAAGTGG ATTTCTAGTA TATGAGTCGT TGAGATATCT AACATAATCA	2040
30	AATCACGACT CAATATTTTC TTACTTTGAG TGC GCGCTTT AACTAAGTTT TCGCGATGCA	2100
	TTGATGTGAC CAATCTCTGA GTGACTTTAT CTCTTCCTTG TAAAATCATC TCTTTAAAAA	2160
35	TATTAGCCCA ATCGCTATTA TGTGTGTTGA AAAATAAATA TGTTCCTTGT TCTTCTTTGT	2220
	TAAATTTTAA TGTCTGTTCT TTAAAAATTA AGTTCAGTT CATAATTCAC CTCTATGAAA	2280
	TATTTTACAA AAGCAAGATA GATTGTGATA ATCCATATTA ATGATAATGA yTCTTATTAT	2340
40	CAACAGAATG CGGGTGTAAG TTTTATGACA AAATATTTTT TTAGCAGTTC TTTTCTACTA	2400
	TTTCTAGGTA ATTGGATTGG ACAAATAGGG CTAAATTGGT TTGTACTTAC CACTTATCAT	2460
	AACGCAGTTT ATCTGGGGAT TGTCATTTTT TGCAGACTTG TACCAATATT ATTACTAAGT	2520
45	GTGTGGGCAG GGGCAATTGC CGATAAATAT GATAAAGGGC GATTGCTGAG AATTACAATT	2580
	TCATCATCAT TTTTAGTAAC TGCAATTTTA TGTGTGCTCA CGTATAGTTc ACTGCAATTC	2640
50	CAATTAGCGT CATTATTATA TATGCGACAT TAAGAGGGAT TTAAAGTGCG GTTGAAACAC	2700
	CTTTAAGACA AGCAATCTTA CCAGATTTAT CAGATAAAAT ATCTACTACA CAAGCTGTmw	2760
	CATTTCATTC ATTCATCATT AATATTTGTC GTTCAATAGG GCCTGCCATT GCTGGTGTCA	2820
55		

	CAGTTTTATT ATGCTTACCA TTACATTTTA AAGTAACTAA AATACCTGAA GaTGCATCAA	2940
	GaTACATGCC GTTAAAAGTT ATTATAGATT ACTTCAAATT ACATATGGAA GGTGACAAA	3000
5	TATTTATAAC ATCATTATTG ATTATGGCGA CAGGTTTTTC ATATACGACA CTTTTACCAG	3060
	TTTTGACAAA CAAAGTATTT CCGGGGAAAT CTGAAATATT TGGTATCGCT ATGACGATGT	3120
	GTGCCATTGG TGGTATTATT GCAACGCTAG TTTTACCTAA AGTACTTAAA TATATTGGTA	3180
10	TGGTAAATAT GTATTATTTA AGTTCATTTT TATTTGGCAT TGCTTTGTTA GGTGTGGTAT	3240
	TTCACAATAT TGTCAATCATG TTCATTTGTA TTACATTGAT TGGGTTATTT AGTCAATGGG	3300
15	CACGTACGAC AAATCGCGTT TATTTTCAAA ATAATGTTAA AGATTATGAA CGTGGTAAAG	3360
	TACTGAGTAT TaTTATGATG GgATAGAGGT ATGaTTCCAT kGGGAAGTCn ATTAATGAGT	3420
	ATATGTGCAG ATGTGTTTGG CATTGTTAGA ACTTTTTCAA TAATGGGAAT AAGTACTATA	3480
20	TGCATTACAA TGGTATTCTA TTTTATAAAT AGAAAGTTGA AGTTAAAGTT GGAGGAAAGT	3540
	AATCATGGTA TATCTTGAAT GGGCAAAGGC AGATAGAAAT ATTCAATATC GTGTAATTAA	3600
	CGCCATTATT AAAGAACGTA TTTACCCCGA GCAAACATTT ATTTGCAAAA AAGGATCTTT	3660
25	AATTGAAATT CAGTATCATA TGCATGTGTT GACTATTGAA GTTGTTAGAA AAAGTGCAAT	3720
	AGAACGCTAT GAGTTTACAG GTGATATTAC TTATTTAAAT AAAGGTGAAA CGTCATTAAT	3780
	TATAACTTTA GAAGGTTTAT TAGATGTGTT GAATCATGAC TTTGATATCC CTATTTCAGA	3840
30	GCGACTACGC GAAGAGTTAA TACACAGTCG AGATAGTTTA GTTGAAACAT ATAAGCAAAT	3900
	GTCTCACAGA CAAACGTTAA TAAGTCmAG TTTTAAATTT TCAAGGTTAC CACAAGATAT	3960
35	TAACTTTTTT TCakGGTtAC AACATGTAAA AGATAGTGAT AAGACAGATG ATTTAACTTA	4020
	TTCTGAGAGT TTGGTACCAG AGGGGCATCC AACACACCCT TTAACCAAAA CGAAATTGCC	4080
	CTTAACATATG GAAGAAGTAC GAGCATATGC ACCTGAGTTT GAAAAAGAAA TCCCTTTGCA	4140
40	AATTATGATG ATTGAAAAAG ACCATGTTGT GTGCACAGCT ATGGATGGTA ATGATCAATT	4200
	TATTATTGAT GAAATAATTC CCGAATACTA CAATCAGATT CGTGTGTTTT TAAAGAGTTT	4260
	AGGTTTGAAA AGTGAAGACT ATAGAGCGAT TTTAGTACAT CCTTGGCAAT ATGATCATA	4320
45	GATAGGGAAA TATTTTGAAG CATGGnTTGC TAAAAAATA TTAATTCCAA CGCCGTTTAC	4380
	AATACTTcCA AAAGCaACTT aTCatTTaGG ACGATGTCTT TAATTGATAA ACcATACCAT	4440
	GTTAAGTTGC CCgTCGATGC aCAAGCAACA AGTGCCGTTA GAACAGTCTC AACTGTGACT	4500
50	ACTGTAGATG GACCAAAGTT AAGTTATGCT TTACAAAACA TGTTGAATCa ATATCCaGGA	4560
	TTTAAAGTTG CTATGGAACC GTTCGGTGAA TATGCAAATG TTGATAAAGA TAGGGCACGT	4620
55		

AGTGCAAGTC TAGTTAATAA AAATCCAATA GATCAAAAAG TTATCGTGGA TAGTTACTTA 4740  
 GAGTGGTTAA ATCAAGGAAT TACTAAAGAA AGTATTACGA CATTATTGA ACGATACGCT 4800  
 5 CAAGCATTAA TCCCGCCTTT AATTGCTTTT ATTCAAAATT ATGGAATTGC TTTAGAAGCA 4860  
 CACATGCAAA ATACAGTAGT GAACTTGCGG CCACATTTTG ACaTTCAATT TTTAGTGAGA 4920  
 GATTTAGGTG GTTCTAGAAT TGATTTAGAA ACATTACAAC ATCGTGATC AGATATTAAA 4980  
 10 ATTACAAATG ATAGTTTAAT AGCTGATTCT ATAGATGCAG TGATTGCAAA ATTCCAACAT 5040  
 GCTGTTATT CAAAATCAAAT GGCAGAATTA ATCCATCATT TTAATCAGTA TGATTGTGTT 5100  
 15 GAAGAAACCG AATTATTAA CATAGTACAG CAAGTAGTAG CGCATGCCAT TAACCCAACA 5160  
 CTACCACATG CAAATGAGTT AAAAGATATT TTGTTTGAC CAACAATTAC TGTCAAAGCG 5220  
 TTGTTAAATA TGAGAATGGA AAATAAGTA AAGCAATATT TAAATATTGA GTTAGATAAT 5280  
 20 CCGATAAAAA AAGAGGTGTA GTACTACATG GCACACGTTA ACATAAATAT ATCGAAGATT 5340  
 AAaTATAACG CCAAAGTACT TCAAACAGTT TTTCAAAGTA AAAATATGCA ATTCACACCA 5400  
 GTAATTAAGT GCATAGCTGG TGACCGTACA ATTGTAGAAA GCTTAAAAGC GTTAGGTATC 5460  
 25 AATCATGTTG CAGAATCCAG ATTGGATAAC ATAATTAGTA TTGCAGATAC AGGATTTAAC 5520  
 A 5521

) INFORMATION FOR SEQ ID NO: 409:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1261 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 409:

40 AGGTGCTCTr GCaGmAGmGm TATTTGACGA TCTTGCTTTC CAAAACACG ATGATGATTT 60  
 TAACATACTG TCTGATTATA TTGAGACACA TGGTGATTTTC aCATTGCCAA TGTCTGTATT 120  
 TGATGATTTA TATGAAGAAT ATACGGAATG GCTAAAATTT TAATATAATT TTTAATAATA 180  
 45 ATAGTTAGAA CCAGGGTGAT GCAATTCGTT ATCCTGGTTT TAATTTAAAA TAACTAAGT 240  
 TGTGACTAAA AATTAATCAA TTATAGTGAA ATATGGTGCG CTATCTTGCA TAAATTGATA 300  
 TGATTAAC TA CACAGAATTT AAAAGTACAT AATACATAAT AAGGAAGTGA TACAATGGAT 360  
 50 GATAAGCAAC ACACATCTTC ATCCGATGAT GAACGCGCTG AAATTGCAAC AAGCAATCAA 420  
 GACCAAGAAA CTAATTCATC GAAACGCGTT CACTTAAAC GTTGGCAATT CATATCAATA 480

CAAAAAATAA GTGGTTTAAA CAAAACGTAT CAAGCAAAC TAAATAAAAT TGAAATGTG 600  
 TATAAAATCT TAAATAGTGA TTATTACAAA AACAGGACT CTGACAAGTT AAGTAAAGCT 660  
 5 GCAATTGATG GCATGGTCAA AGAATTAAAA GATCCTTATT CTGAATATTT AACAAAAGAA 720  
 CAAACGAAAT CCTTTAATGA AGGTGTTTCA GGTGATTTTG TAGGTATTGG TGCAGAAATG 780  
 CAAAAGAAAA ATGATCAAAT TATGGTTACT AGTCCTATGA AGGGATCTCC AGCAGAACGT 840  
 10 GCTGGCATTG GTCCTAAAGA TGTCATTACT AAAGTAAATG GAAAATCAAT TAAAGGTAAA 900  
 GCATTAGATG AAGTTGTCAA AGATGTTTCGT GGTAAAGAAA AACTGGAAGT CACTTTAACT 960  
 GTTCAACGAG GTAGTGAAGA AAAAGACGTT AAGATTAAAC GTGAAAAAAT TCATGTTAAA 1020  
 15 AGTGTGTGAGT ATAAGAAAAA AGGTAAAGTT GGAGTTATTA CTATTAATAA ATTCCaGAAT 1080  
 GATACATCAG GTGAATTGAA AGATGCAGTT CTAAAAGCTC ACAAAGATGG TTTGAAAAAG 1140  
 ATTGTTTTAG ATTTAAGAAA TAATCCAGGT GGACTACTAG ATGAAGCTGT TAAAATGGCA 1200  
 20 AATATTTTGA TCGATAAAGG AAAAAGTGTG GTTAACTAG AaAAAGGTAA AGATACTGAA 1260  
 G 1261

25 (2) INFORMATION FOR SEQ ID NO: 410:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2488 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 410:

35 AAATATATTG AAnAGAnAAT TACTAAGATT AAATCnTCTT AAAATATCCC TGAAATAACG 60  
 TCCTAAAGAT TAAAGGAAAG AGGTTATAAG TTATGCCAAA ATTAATTTTA TGTCGTCATG 120  
 40 GACAAAGCGA GTGGAATGCT AAAAAGTTAT TTACTGGATG GGAAGATGTT AATTTATCTG 180  
 AACAAAGTAT TAATGAAGCG ACTAGAGCAG GTGAAAAAGT AAGAGAAAAT AACATTGCCA 240  
 TCGATGTAGC TTTTACATCG TTATTAACAC GTGCTTTAGA TACAACGCAT TATATTTTAA 300  
 45 CTGAATCTAA ACAACAATGG ATTCCTGTAT ATAAAAGCTG GCGTTTAAAT GAACGCCACT 360  
 ATGGTGGATT GCAAGGCTTA AATAAAGATG ATGCTAGAAA AGAATTTGGA GAAGAACAAG 420  
 TACATATTTG GCGTCGTTCT TATGATGTGA AACCACCTGC TGAAACCGAA GAACAACGTG 480  
 50 AAGCTTACTT AGCTGATCGT CGATATAATC ATTTAGATAA ACGTATGATG CCTTATTCTG 540  
 AAAGTCTGAA AGATACTTTA GTTCGAGTGA TACCATTTTG GACAGATCAT ATTTACAAT 600

55

	TTAAATATCT TGAAGATGTG TCAGATGAAG ATATCATTAA TTATGAAATT AAAACAGGTG	720
	CACCGCTTGT TTATGAATTA ACGGATGATT TAGAAGTTAT AGATAAATAC TACTTATAAA	780
5	AaAAGAGCTG CATGTACACA AGGAGTGAGT GTATATGcAG CTCTTAAAtA TGTGAAGTAA	840
	TGTAAGGAAA TAGTTAAGTA TAGAGTTTAT ATTAACGAGC TAGGGATACT CGAAAATATA	900
	GTTAGACATA CAATATAGTC AAATTAAAAC AATTATTTTCG CTCTTTTATG TTGCTTAATA	960
10	ATCTTTAAAG CACGCTTTCT TGTTTTAATG TTAGGGCTAT TTAAATTACG ACGAGCAGTC	1020
	TGTAAATCTA ATTTTCATCTC TATCCCTCCT TGTAATATA TTATGACCGA TAACTACTCA	1080
15	TATGTAAATA GTAATGATTA CGTTTTAAAG AAATTGTAAT AAAGTCGTGC TAATTTTTTG	1140
	GAAAATGGGT ATAATTACCG GATATCTAAA AATGTGTGTC GTTTTTTAGA TGGTGAGGGG	1200
	GAAGCTTTAA ATGTGGAAGA AACAAAAATT AACGATGATT ATTACTATGC TGATGGGTGG	1260
20	ATTTTTTGGA TTATTAAATG AAACACTATT AGTGACGGCT TTACCAAGTA TTATGAAAGA	1320
	TTTTGAAATT TCATATACAC AAGTTCAATG GCTGACAACA GCTTTTTTAT TGACTAATGG	1380
	GATTGTTATT CCTTGTCCG CGCTTGTTAT ACAACGTTAT ACAACAAGAC AAGTGTTTTT	1440
25	AGTGGGTATT TCTATCTTTT TCTTAGGTAC ATTACTCGGC GGCTTGAGTC CGCACTTTGC	1500
	AACATTATTA GTTGCTAGAA TTATTCAGGC GTTAGGCGCA GGTATTATGA TGCCATTGAT	1560
	GATGACAACG ATTTTGATG TTTTCCAACC ACATGAACGC GGTAATATA TGGGGATATT	1620
30	TGGTTTGTA ATTGGTTTAG CACCAGCTAT TGGACCTACT CTTTCAGGTT ACCTTGTTGA	1680
	ATATTTTAAC TGGAGATCGC TTTTCCATGT TGTCGCTCCA ATTGCAGCTG TGACATTTTT	1740
35	AATTGGaTTT AAAAcGATAA AAAATGTTGG AACTACAATT AAaGTACCTA TTGATTTTAT	1800
	TTCTGTcATT TTTTCTGTAC TAGGTTCGG cGGGTtATTG tATGGAACGA GTTCaATTtC	1860
	AGAAAAAGGT TTTGATAATC CTAcGtATTA GTATCTATGA TTGGAGGCGT TGTTTTAGTC	1920
40	GCATTATTTG TAwtACGTCA ATATCGGCTA TCAACACCAT TATTaAATTT TGCTGTATTT	1980
	AAAAATAAAC AATTTACAGT TGGTATCATT ATTATGGGTG TCaCAATGGT ATCGATGATT	2040
	GGTTCGGAAA CGATTTTACC TATCTTTGTG CAAAATTTAT TGCATCGTTC AGCTTTAGAT	2100
45	TCTGGATTAA CTTTATTACC AGGAGCAATT GTTATGGCAT TTATGTCGAT GACTTCGGGT	2160
	GCTTTATATG AAAAGTTTGG TCCTAGAAAT CTTGCTTTAG TAGGTATGGC GATTGTTGTT	2220
	ATTACTACGG CTTATTTTGT TGTAATGGAT GAACAAACAT CAACAATCAT GTTGGAACA	2280
50	GTTTATGCGA TTCGAATGGT TGGTATCGCG TTAGGATTAA TTCCAGTAAT GACCCATACG	2340
	ATGAATCAAT TAAAGCCAGA AATGAATGCA CATGGTTCAT CTATGACAAA CACAGTACAA	2400
55		

AACTTTTCTC CAACTATGTC AGACTATA

2488

(2) INFORMATION FOR SEQ ID NO: 411:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1105 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 411:

15	TTTTACAATT TCAGATATCT CTAAACACA TACCGATCCA ATGGCCAAAA TTATTCGTCA	60
	AAAATTGAAA AAACTAGGAA TTCGTAAAGG GATTCCAGTT GTATTTTCAG ATGAAAGTCC	120
	AATTGTCATA AGAGAAGATG TAAAAGATAT AGTTGGAGAT AAAAATGCTA TCAATCGAAA	180
20	AGGACAAATG CCACCTTCTT CAAATGCCTT TGTGCCAAGT GTTGTGGAT TAATTTGTGC	240
	AAGTTATGTG GTGAATGACG TATTAAAAGA TATTCCAGTT CGTCGCATTA AAGACAAAGG	300
	GCAATAATTC ATTTTGAAAG GGATAATTTT CAACGTAAGG CAAGTGTAAC CACACCATAA	360
25	AACTAAACT GACTAGTTCG CAAGCATAGT AGAATATGCC TCGTGACTA GTCAGTTTGT	420
	ATTTGATTAT AGCTAAAATA TCAAAGGTAA TCGATTTAAT CTTATTTTTA CGGCCTTTGT	480
	TTAAGTAAGT TGTTATATAT TTCTTTGAGT TGTGTTCAC TTTTAGATGT CGTTTTTGGT	540
30	TCGTAATAAA TTTTGTTTTT TAGTTTATCA GGTAAATATT GCTGTGAAAC ATAGCCATTA	600
	ACATATTGGT GTGGATATTT GTAACCAATA GATCGGCCTA GATCTTTAGC ACCTTGATAA	660
35	TGTCCATCTT TTAAATGGTT TGGAATTTGG CCCACATGAC CGTTTCTAAT ATCGGACAAT	720
	GCACTATCAA TGGCACTCAT TGCTGAATTT GACTTAGGTG ATAAGCATAG TTCGATTACT	780
	GCTTGGCTTA GTGGAATTCT AGCTTCTGGT AAACCTAGAC GTTCTGCTGA TTCAATAGCA	840
40	GCAAGTGTTC TCTGACCAGC ATTAGGTGAG GCTAAGCCTA TATCCTCATA GCTAATTACA	900
	AGTAATCGTC GAACTATTGT AGGTAAATCT CCAGCTTCAA TTAATCGTGC TAAATAATGT	960
	AAAGCGGCAT TGACGTCGCT ACCACGGATA GATTTTGGGA AAGCGCTCAT AACATCGTAA	1020
45	TGCATGTCAC CATCCTTGTC ACTTACAAAT GCACCTTTTT GTAAACAGTC TTTAGCATCT	1080
	TGCAATGTAA CATGTCGATA ACCGT	1105

(2) INFORMATION FOR SEQ ID NO: 412:

50

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 579 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 412:

5 TAACGTCGAA CTTGAGCTGT TACGTTATGA CTCATAATTA TTTTAGCATA GTCATTTAAA 60  
TAAACTTCTG TTCTGTCTGT TGGATAAGCA AATTCAAGCA ATTGACTGTA ACTATCAITTA 120  
ATGACTTCTT GATCAACATG ACTATCAAAA TATACAGCAT AATAATAAGT ACCATCAACC 180  
10 ATATATAACA AATCTTCAAA CTCTGTAGTT ATTGGATTGC TATGATATGC ATAATTAATA 240  
ACATCTTCTA AATCGTTAAA TTTCACAATG ATTGTTCTTG TATTTTTACG TGCTGAAGAC 300  
TTTTGACGTT TAGAACCTTG AGCTTCTTTT TCTTTTGTTT GTTGCTCGAA TAATTCTTCT 360  
15 AATTGATCTT CACCTTCTAA TGTTTGAGCT AACAAATCTT GAACTTGTTT ATCAAATKGA 420  
TCAGTTGCAT CATCATCAGA CATATTCATC ATATCTTCAT TTTTAGATTT AGAAATTGTG 480  
ACTTCGACAC CTTTTTCAAA GGCATGTACT TGAATCCATA ATGGACCTnC TACAACAAAA 540  
20 TCTTCTACTT CGTTAATTTT ATCCATCATT GAnCAAAAG 579

## (2) INFORMATION FOR SEQ ID NO: 413:

(i) SEQUENCE CHARACTERISTICS:  
25 (A) LENGTH: 1342 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 413:

GTGGTGAAAC TTCTTGCTTT GTAATTTTAT AAAGTGGATC AATATCACTC TGACTTACAT 60  
35 CAGGTTGAAC TGTCATCTTT TTAGTTACTT TGTTTTCAAG CGTAATATTT GCTAATTCTA 120  
GGtTTTACAG AATAATGAAT GTCATCAGCT AmCCCTTTmm CtTGATACTy CTCTTGATAT 180  
TTTCCAGCTT TAGCATTTGA TAAATCAATC ACTACTCTTA AATCTTCTGG ATTTTCAATT 240  
40 TTTATTATCT TTGATTGTGG TCCTGAAATA GTCACATTAA CTGTTTCAGG CGCTTTTGTT 300  
AAATGCAAAT CTTTAGTGTT ATAAAGAATT TCAACGGGTA CATCTTGAAT CGTTTACTA 360  
GACTTTTGAC CAAGATTACC AGTGTTAAAG ATATTTCCAA AACATTGTT AACAGATAAA 420  
45 AAGAAAAACA ATGCCAAAAG AAAGGCAATA AATCTCAAGC CCCATTTACT TTCTAGCATA 480  
TTATTTTACA CTTTCTTTT GAAAGCGTGT GCCAAACCAA TGTTAGCAA GCAACTCTTC 540  
AAAAATTTG TTTGAAATGT CTCGTCGTAA TTTTCCATCA AATGTTACCG AAATATCACC 600  
50 AGTTTCTTCA GATACAATAA CGGTAAATGC ATCAGATACT TCTGAAATAC CAACCGCAGC 660

55

TGCTGCTGCT GCAATCTTCG TGCCTTGAAT AATCATTGCA CCATCATGTA AAGGTGTGTT 780  
 AGGTATAAAG ACATTAATTA AAAGTTCTTG CGAAATATTT GAATCCATTG CAATACCTGT 840  
 5 TTCAATATAA TCTTGAAGAC CTGTTTCTTT TTCAAAGACA ATTAATGCAC CTATACGTCT 900  
 TTTAGCCATA TATTGCACAG CCTTTGAAAC CGATTGAATC AATTCTCTT CATCTTTACT 960  
 ATACGTATTA GAAGTATAGC GTTTTAAAAA GCTACCTCTA CCAAGTTGTT CTAACGCACG 1020  
 10 TCTAATTTCT GGTGGAATA TTACTATTAA AGCTAATACC CCCCATTGAA TAACGATATC 1080  
 GAATAATTTA GATGTGCGAG TCAAGTTCAA TATCATACTT ATCTGCTGAC CAATAACAAT 1140  
 15 TACTAATATC CCTTTAAGTA ATTGTATCGC TTTAGTCCCC TTAAAGACCG TGATGAGAAG 1200  
 ATAAAGTACA TACCAAATA TCAGTAAATC AAGGATACTC GTTACAATTT TTAACGTACT 1260  
 GAGGTTTTGA AAAAAGTTGG AAAAATCCAT AACATCTCCT CCGGGTAATA TTTTCCATA 1320  
 20 ATACCCATTA TACCAATCAT TT 1342

## (2) INFORMATION FOR SEQ ID NO: 414:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1073 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 414:

CTGTATAAAG ATGGAGGTGT TTTGCATGGT AAAACGTACT TATCAACCAA ATAAACGTAA 60  
 ACATAGTAAA GTTCATGGTT TCAGAAAACG CATGAGCACA AAAAATGGCC GTAAAGTTTT 120  
 35 AGCGCGCCGT CGTCGTAAAG GCCGTAAAGT TTTATCTGCA TAAGATCACT GACTCATCAG 180  
 TGATCTKTTT TTTCGTTTAA ATTAAGAATA AATAGAAATT TATGTTATAA GCTCAATAGA 240  
 AGTTTAAATA TAGCTTCAnA TAAAAACGAT AnATAAGCGA GTGATGTTAT TGGAAAAAGC 300  
 40 TTACCGAATT AAAAAGAATG CAGATTTTCA GAGAATATAT AAAAAAGGTC ATTCTGTAGC 360  
 CAACAGACAA TTTGTTGTAT ACACTTGTA TAATAAAGAA ATAGACCATT TTCGCTTAGG 420  
 TATTAGTGTT TCTAAAAAAC TAGGTAATGC AGTGTTAAGA AACAAGATTA AnAGAGCAAT 480  
 45 ACGTGAAAAT TTCAAAGTAC ATAAGTCGCA TATATTGGCC AAAGATATTA TTGTAATAGC 540  
 AAGACAGCCA GCTAAAGATA TGACGACTTT ACAAATACAG AATAGTCTTG AGCACGTACT 600  
 50 TAAAATTGCC AAAGTTTTTA ATAAAnAGAT TAAGTAAGGn TAGGGTAGGG GAAGGAAAAC 660  
 ATTAACCACT CAACACATCC CGAAGTCTTA CCTCAGACAA ACGTAAGACT GACCTTAGGG 720

55



TTTAGATACA ATTACGAGTA TTTCAACACC AATGGGTGAA GGGGCAATTG GAATTGTTTCG 840  
 ATTGTCTGGA CCGCAAGCCG TTGAAATTGC TGACAAATTA TATAAAGGAA AACATCTTTT 900  
 5 AAATGATGTT CCATCACATA CGATTAATTA CGGTCATATT ATTGATCCAG tCTAAAGAAG 960  
 TGGTTGAAGA AGTTATGGTG TCTGTGTtAA GAGCGCCAAr AACATTTACA CGCGArGATA 1020  
 TTATAGAgAT TAATTGTCAT GGTGGtATTk TAACGATTAA TAGAGTGCTG GGA 1073  
 10

(2) INFORMATION FOR SEQ ID NO: 415:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3176 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 415:

CTTTACCAAT GCCAAATCCG AAGTAAAGTA TAGCAATAAA GATTACTAAT ACAATTCTGT 60  
 AAATGGCAAA TGGAATTAGT TTGATTTTGT TAATTAGATG CAAGAATGTT TTGATTGCAA 120  
 25 TTAGTCCAAC AGTAAATGCA GCTAAAAGC CTAAATATA AAAAGGTATA TCAGCAATCT 180  
 GAATATCTTG ATAATGTTTT AATAAGATA AACCCTAGC TGCTAACATA ATTGGAACAG 240  
 CCATAATAAA TGTAAGTCC GATGCTGCTT TATGATTTAA TTTCATTAAT ACCCCAGTTG 300  
 30 AAATTGTTGA GCCTGAACGG CTGAAACCAG GCCACATAGC TACTGCTTGA gAAATACCAA 360  
 TTACAAATGC TTGGAATAA CTGATTGAT CTACTGTTTG TGGGTTTTTA ACTTTAGCTG 420  
 AGTATTTATC AGCAATAATC ATATAGATAG CACCTACGAA TAAGCCAATC ATAACAGTTG 480  
 35 GCACACTAAA TAAATGTTCT TCGATGAAAT CATCAAATAG TAAGCCTAAA ATACCTGCTG 540  
 GCACCATACC CACTAATACA TGTAATAAAT TTAAACGTCT TGGCTTTGAA CGTCTTTGTT 600  
 40 GATCGTTATC TCCTTCAACA TGTTTGTGTT TACCAATATG TAAAATCTCT AAGAAGCGTT 660  
 CGCGGAACAC CCATGCTGCT GCAAAGACGG ATCCTAATTG GATGACGATT TTAAATGTAA 720  
 ATGCTGACTG AGAACCTAAA AATTCAGATG ATTTTAACCA CATATCATCA ACTAGGATCA 780  
 45 TATGTCCAGT AGAGGAAACA GGTGCAAAT CTGTTAATCC TTCGACGACC CCTAAGATAA 840  
 TACCTTTTaT TAATTCAATG ATAAACATAA TGTACCCACT TTCATTACTC AATTTAATTT 900  
 ATTTAAATAT CAAAATTACC ATATCATGAT AGCATATTCA TTTAAAGACA TGCTAGTTAT 960  
 50 AGTTATAATA CTAGACTAAA GATGTATATA TTCATTTTCT TTTACATGTA AACTACAAT 1020  
 ATTTTATTGA GCTATTTAAT TTGATTTTAA GGAAACCTT TTATAATAGG TTTAGGTGAT 1080  
 55

	TTCTTGGTCA GTACTGGTCT CGGCATAATC GTTATAACGC AAAATATTTT AATAGCAGAT	1200
	TTTTTAGCTA AAATTATAAG ACATCAATTT CAAGGTTTAT GGATTGTATT ATTTATTTTA	1260
5	TTAGGTGTTT TACTTTTAAG AGCAACTGTG CAATTTCTAA ATCAATGGTT AGGTGATACA	1320
	TTAGCATTTA AAGTTAAGCA TATGCTTAGA CAGCGGGTTA TTTATAAAAA TAATGGTCAT	1380
	CCAATCGGTG AACAAATGAC TATACTCACA GAAAACATTG ATGGTCTAGC ACCTTTTTAT	1440
10	AAGAGTTATT TGCCTCAAGT GTTCAAATCA ATGATGGTTC CGCTCATCAT AATCATTGCA	1500
	ATGTTTTTCA TCCATTTCAA TACCGCAITTA ATTATGTTAA TAACTGCACC ATTTATTCTT	1560
15	TTGTTTTATA TTATTTTCGG TTTGAAAACG CGAGATGAGT CAAAAGATCA AATGACTTAT	1620
	TTGAATCAAT TTAGTCAACG GTTTTTAAAT ATTGCTAAAG GTTTAGTGAC GTTAAAGCTA	1680
	TTTAATCGTA CAGAGCAAAC AGAGAAGCaT ATTTaCGACG ATAGTACTCa GTTTAGAACT	1740
20	TTAACAATGC GCATTTTaCG CAGTGCTTTT TTATCGGGAT TAATGCTCGA ATTTATAAGT	1800
	ATGTTAGGTA TTGGATTGGT TGCATTGGAA GCAACGCTAA GCTTAGTAGT ATTTCATAAT	1860
	ATTGATTTTA AAATGCGGC AATTGCGATT ATTTTAGCGC CTGAATTTTA TAATGCAATT	1920
25	AAGGACTTAG GGCAAGCGTT CCATACTGGA AAACAAAGTG AAGGTGCCAG TGACGTTGTG	1980
	TTTGAGTTTT TAGAACAACC GAACTATAAT AATGAATTC TATTAAAGTA TGAGGAAAC	2040
	CAAAAGCCAT TTATTCAGTT AACAGACATA TCATTTGAT ATGATGATTC TGATAGATTG	2100
30	GTATTAAATG ATTTAAATTT GGAAATATTT AAAGGTGATC AAATTGCACT TGTAGGTCCA	2160
	AGCGGGGCAG GTAAATCCAC TTTGACACAT CTTATTGCAG GTGTTTATCA GCCAACAATA	2220
	GGTACTATAA GTACAAACCA GCGTGATTTA AATATAGGAA TACTTAGTCA ACAGCCATAT	2280
35	ATTTTCAGTG CTTCTATAAA AGAGAATATT ACGATGTTTA AAGATATAGA AAATAATACT	2340
	ATTGAAGAAG TGCTAGACGA AGTAGGTTTA TTAGACAAAG TGCAATCTTT CACAAAAGGC	2400
40	ATTAACACAA TAATAGGTGA AGGAGGCGAA ATGTTATCTG GTGGACAGAT GAGACGCATA	2460
	GAACTTTGCC GTCTTTTAGT TATGAAGCCA GATCTCGTTA TATTTGATGA GCCTGCAACT	2520
	GGTTTAGATA TTCAAACAGA ACACATGATT CAGAACGTTT TGTTCACAA TTTTAAAGAT	2580
45	ACAACGATGA TTGTCATTGC ACATAGAGAT AATACAATTC GCCATTTACA ACGACGCTTG	2640
	TATATAGAAA ATGGAAGACT GATTGCTGAT GATCGCAATA TTTCAGTAAA TATAACAGAA	2700
	AATGGTGATG ACTTATGAAA ACACGACTAA AATTTCAAGT AGATAAGGAT TTATTGTTAG	2760
50	CTATAGTTGT TGGTGTGTTGT GGAAGTTTAG TTGCGCTCGC CATGTTTTTC TTAAGTGGTT	2820
	ATATGGTGAC ACAAAGTGCA CTTGGTGCGC CACTATACGC TCTGATGATT TTAGTCGTTA	2880

55

ATAAAGCTAC ATTTACAATG CTACGTGATA TTCGGGTACA GTTTTTTCGGT AAATTAGTAA 3000  
 ATGTCATTCC TAATGTTTAC CGTAAACTGA GTTCTAGTGA TTTAATTTC ACGTATGATTA 3060  
 5 GTCGTGTTGA GGCATTACAA AATATAkATT TACGTGTTTA TTATCCACCA GTCGTCATCG 3120  
 GTTTGACAGC GCTAGTTACA GTCATAGTTT TGGCGTTCAT TTCAATCGGC CATGCG 3176

(2) INFORMATION FOR SEQ ID NO: 416:

10

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2109 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 416:

20 TCTTTATTTT AATTTC CAAT TGAATTTTTT TATTATTTAC GCATAGCTCT TAAAATTAAC 60  
 GTTACGATTG CAATTAAGAT AATTGAACCA ATTAATGCTG GCAAGATGTA AATACtTCCT 120  
 AATTCAGGAC CCCATTGTCC GAATAGTTTG CCACCTACCC ATGATCCAAT AATACCTGCG 180  
 25 ATAATATTGC CTAAAATACC ACCTGGGATA TCTTTACCCA TAATAGCACC AGCAGCCCAT 240  
 CCAATTAAGC CACCGACAAT TAACATTCCA ATAAATCCAA ACATAATTTT CAGTCTCCTT 300  
 TTTCTATTTA TTTTGC GTTA TTCTAAGTAG TACCCCTTAT TTACAATTCT AAAACAAATT 360  
 30 CAAATTATTT TTATCCAAAT ATTTTAAAAA GTAGTAATTG AATATCAATT TTATTCAATG 420  
 TAGCTATCGT TATTTAAAGT CTCTGTACCG ATAATATCAT ATACATTTAC ATTATTTTTT 480  
 CTGCCGAATT CATAGCTTGA TTATTTTATG TTATAGGACT AgAATATACA CATATTATTA 540  
 35 GAGCATCTTT GAATTTTAAA TCAAGAAGCG AGGTTAATGA ACAATGAATA TGCATATTTT 600  
 ATATaACTTA CGAACTAAAC ATAATTTAGA AATTGACGAA TTAGCACAGC AATTAAATGA 660  
 GAAATATGGT ACTAAATATG AAGCACATCA AATTTGGGAA TGGGAGAATC ATCACCATGA 720  
 40 ACCTAAATTT AAAGATGCCA TGCATTTAGC TGACTtCTTT GATGCACCAT ATGAAATGTT 780  
 TTTAGAAAGT AAGGTTAAAG AATATCAGAA ACATTTAGAA GAAGTCGATA TTCGCATGGA 840  
 TAAATAGATG CAAATAAACC CTCACAACAC GTTTGGCATA TATCCTTTCA AATCTATACT 900  
 GGATATATTA CATTACGTTG TGAGGGTATT TTATTAATTA ATATGAATTA AGACATTTTA 960  
 CAAGCGTTAA TGCAACGAAT CTTTTTAGTG ATCTTGCTCA CTCTTTAATA CTTTACCGTT 1020  
 50 CTTAGCATCA ACAGTAACTT CTGTGTTTTT ATTACCTTTT TTCAAATCGA TATTGTAAAC 1080  
 AAGTTTGCCA TCATCTTTTT CAAGTGACCA TTCTTTAATA TCACCATCAA ATTCTTTTTG 1140

55

ATTCATTGTA TCTTCTTTTT CAGTCTTTTT GTTAATCACT TTTTATTTT TATCAGCAAC 1260  
 AAGTACTTCT GACTCTTCAC CAGATTTTTG TTGCGTCACT TTATAAGCCC ATTCACCATT 1320  
 5 AGAATTTTCA AATGAAATTC CTTTCAACTT TTGGCCTTTG TAAGTTTCTT CAGCTTTTTT 1380  
 CACAGCATCT TCTGGGCTTG TTTTAACATC TTTTAAAGCA ATAACATCTT TTGTGTATT 1440  
 AGTGTCTTGA TTAGTATTG ACTCTGTTGA TTTGTTTCA TCTTTTGGAG TATCATTGCC 1500  
 10 ACATGCAGTA AGCACCACCG CTGACATTGA TAACACTGCT AATGATTTTA ATTCATAAT 1560  
 ATCACTCTCT CTTCTATTTT TTGAACTCA TAACAAAAGC TTATATGCTA TATAGATTGT 1620  
 ATTACCCCTT GTTTTTAATT TTATTCATAA TTATTACAAA TATTTTAAA TTAATCGTCA 1680  
 15 TCGGTTACTT TCGTTCGTAC TCCTTTTATA AATGAACCAT GTAATATAAG CATGCTATTA 1740  
 TCGACTCAGT TTGTCTAATG CTTTCTTTGG TACTTCTTCC TTTTCAACTT CTTCAAAAGT 1800  
 20 TTCTACATGA TGACCTTTAT GTGTAATTTT CAAGTATCTA TCGGGTTTAA CATCAAATGT 1860  
 AGCAGTATAC ATTAATTGCG TCTCTTTCCC TTCTTTATTA AAAACACTTT GACTATAACT 1920  
 TCGGAGTTGA TCATCCATTC TAGTAGACAC CTCTGTCGTT TTAACATAAG AATCATCCTT 1980  
 25 CTTTACTAAT GGGTTAAATT GATCTGTTAT ACCATGGGAG TCTATTGTTT TTAAATGAA 2040  
 TAAAGCAGCA TAAGCGCCAA TGATAGTAAG TACAAGATA GTKATTGTTT TTAAATCGT 2100  
 TTCAAAAGA 2109

30 (2) INFORMATION FOR SEQ ID NO: 417:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 813 base pairs  
 (B) TYPE: nucleic acid  
 35 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 417:

GTTACATAAA TGAAAACAAA AAAGATAATT TTAGTCTTA TGCTACACCA GAACATAATT 60  
 ATCAATTGTTG TGGTGCTATG ATAGAAAGTG AAAAATTAAG CGAGTTACTA AAGCCAGCCA 120  
 45 ATCAGTTAAA ATCACCAGAT GATATAAAAA AAGAACTAAA TAAAAAGAAG AGTCACTAAA 180  
 GTTAGGAGTT ACTTTAATGT CCAAAAAACA TGTTTTTATA ATTATTGGTG TCATATTGTG 240  
 TATATGTACA GTTTCTACGG TCATGCATTT TAAAATGAAA TATGATGAAA AAGAAAAACA 300  
 50 AAAAGCGATT TACTACAAAG AACACAAGA ACGTATTACA CTCTATCTTA AGCATAATAC 360  
 TAAAGAAACG AACACGATTA AATCTGTACA TTTCACAAAC TTGGAACAA GTCCTATGGG 420

55

ATCGCCTGAA CATAATTATC AATTGGTGG CGCTATGATA AAAAGTGAAG GAGTAGATAA 540  
 ATTATTAAAA CCAGCACATG AAAGAAAATC ACCAGAAAAA ATCAAAGAAG AATTAGATAA 600  
 5 AAAAGAAGGC CACTAGGGTC TTCTTTATTT TTGATTTAAT CTTCCAATAA TCTATGTCAT 660  
 TGCTATCGAA GGTGTTTCGC AATTAATATA AATCACTTCA TCATCACCAA TACTTCCCCA 720  
 GTTTGTACA GTACATTAAC ACAAACGAAC CACGTTAATT TAAATGGAWT AaTAGTTTGG 780  
 10 CCATTATAAG AACAATATAT ATCGAnTAAC AAT 813

(2) INFORMATION FOR SEQ ID NO: 418:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 640 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 418:

AAAAGCAATC GTTGGTGGTG CTAAATTCAT CGGCAACTCA TATGTAAAAG CTGGTCAAAA 60  
 25 TACACTTTAC AAAATGAGAT GGAATCCTGC ACATCCAGGA ACACACCAAT ATGCTACAGA 120  
 TGTAGATTGG GCTAACATCA ATGCTAAAAT CATCAAAGGC TACTATGATA AAATTGGCGA 180  
 AGTCGGCAAA TACTTCGACA TCCCACAATA TAAATAAGCA ACATGAACAT AGGATCAAAA 240  
 30 GTCATCCCCC ACTATCAATC ATGGGGGATG ACCTTTGATC CCTTTTTTAT ACATACACAA 300  
 GCAAAAATAG CGGTGATTGT TTACCATCAA TTTTAACAAT CACCGCTACT TTTGCTTGTA 360  
 ATTCATGATT CAATTTTTGT TGTGTGCACA ACGACACTAA ATTATGTGTT TGCTATTGTC 420  
 35 GTGTTACAAC GATATGCGTC GTTGATTTAA CTTATCAAGT AATTGATTTA AATTGTCTAA 480  
 TTCGACTTCC GATAAACATT GACATCTTGC TTCAATCAAT TCGCAACGTG CATTATTTAT 540  
 TTGTGAAATT AATGTACGTG CTTGATCAGT CAAAATTAAT TCTTTACATC TTAAATCTTC 600  
 40 GCTAGATTGT TGACTTGTGA TGTACCCnnt CAAAACATAAT 640

(2) INFORMATION FOR SEQ ID NO: 419:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1159 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 419:

55

TTCTTGATA CGTAATAAAT TACCTTTTGT CATAGGATTA AACATAGATC CAGGTAAAAC 120  
 ATAAACATTT CTATTTTGCT CTAATGCAAA ATCTATCGTG ATGTGACTGC CACTTTGTTC 180  
 5 CTTAGCCTCA GTAATTAAAA CCCCTTTTGA CAAACCGCTG ATAATTCTAT TGCCTCAGG 240  
 AAATCTATAT TTAGCGATTG GTGTATGTGG TGGATATTCA GATATAACTA AACCTTTTTTC 300  
 TTCTATTTTA TTTCTTAATG CTAATGTACT TTTGGGATAA TGTGTTTGAT GGCCAAAGGC 360  
 10 TAAAACTGCA ATTGTAGGGA GATTGTATTT TAAAGCTATT TGATGTGCCA TTGCATCAGC 420  
 TCCTTGAGCA AGGCCGGA AATTGTGTA ATATTTGCTT TTATCATTTG ATAATAAAAA 480  
 TTCTAAAGAC TGTGGGTAT AACTGTAGA ATCTCTTGCA CCTACTACTG CCAAATGATG 540  
 15 CATATTATTT ATTAATTTGA TGTCCCTTT ATAGAAAAGA AGTAATGGAA ATTGATATAT 600  
 TTCTTTTAAT AGCACTGGGT ATTCAGAATC CATATATGTA ATGTAACCTA CTTTAAATTT 660  
 CTGCAGTTCT TTAATAATAT CGTTATGATC AAGTTTTATA AAAGCATAGT ATTTACGTAA 720  
 20 TAAATGAACA TTTTCTTCCC TATTCACCCA TTCACATAAA TAACTATCTT TTTTCTTCC 780  
 CTCCTCTTTA ATTACATTAG GATATGCCAT TAAAAATTGA TGAATTGTGTT TAGTCGAAAA 840  
 25 GTGTGCCCAG TATAACTTAA GCAAAAATAG TTTAATCAAT AAATCAACTC CTTTTTGTA 900  
 TCATACAAAA TCATATTCTA TTTTGTGTTT ACATTTCTAA TACAAAAACA TTGTCGATGT 960  
 AATGTTATTT TAAGGAGTAA AATACTGAC TAAAAAGTG AAAAGTATGT TGGAAAGAAT 1020  
 30 TTAAATTTT AATTCCAAC ATACTTTATA ATTAAACCTT ATAAATAAGT TTTGCAAATT 1080  
 TATATAGAAT TGGTCTTACT GGTGTGATGA AATCACC AAT TAATCTTCA ACATGTGCAT 1140  
 TAAACCCTT TTTAAATTG 1159

35 (2) INFORMATION FOR SEQ ID NO: 420:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1879 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 420:

TCAATCAGCG ACTACAAGAA GTGCGGGTCT TCAACAATT GATGTGACAA CACTAAGTGA 60  
 CCCCACTAAT ATTATCATGG GTATTTTAAT GTTTATAGGA TCTTCGCCAA GTTCGGTTGG 120  
 50 TGGCGGTATT CGTACAACAA CTTTCGCTAT TTTAATTTTG TTTTAAATTA ACTTTAGTAA 180  
 TAATGCCGAT AAAACATCCA TTAAAGTTTA CAATAGAGAA GTACACATTA TGGATATTCA 240

55

ATCAGCTACT GAAAATGGTA AGCTTACATT TTTACAAGTA TTTTGAAG TCATGTCTGC 360  
 ATTTGGAAGT TGTGGACTAT CGCTTGGTGT CACAAGTGAT ATTAGTGATA TTTCTAAGGT 420  
 5 CGTACTAATG ATACTCATGT TTATAGGACG TGTGGCTTA ATATCATTTA TCATTATGAT 480  
 AGCAGGACGT CGAGAACCAG ATAAATTCCA TTATCCAAAA GAACGTATTC AAATAGGATA 540  
 ATATAATAGC AATCTAAGTT TAGTTAATGT AGATTTTAAC TGGAACTTAG ATTGCTTTTT 600  
 10 TAGTTTGAT TTTTAACTTA TTTTATAAGA CGATTGGTTT CGAAAATGGT AACTAGTAA 660  
 CAATGAGAGG TGTAACATGA TGGAAAAAAA TGAAACATT AATGTAGAGA TTTTAACTAC 720  
 GTCAGATATG CATAGTCATT TCTTAAATGG TGATTATGGT TCAAATATTT ATAGAGCTGG 780  
 15 TACTTATGTT AACCAAGTAA GAGCACAAAA TCATCGCGTC ATTTTATTAG ATAGTGGCGG 840  
 AAGTTTAGCT GGCTCGTTAG CGGCCTATTA TTATGCTATT GTTGCACCTT ATAAACGACA 900  
 20 TCCAATGATA AAGTTAATGA ACAGAATGCA TTATGATGCT AGCGGTGTGA GTCCAAGTGA 960  
 ATTCAGTTT GGTTCATCAT TTTTAACTCG TTCAATTGCT TTGGCAGGTT TTCCATGGTT 1020  
 ATCAGCAAAT ATTGAATACA ATGTTACTAA GGAGCCTTAT TTTTCAACTC CATATTGTAT 1080  
 25 TAAACATTTT GGTGACTTAA AAATTGCTAT CGTAGGCGTC ACAGCAGATG GTTTAATGGA 1140  
 AAATGAGTAT TCTGAAATGG AGCAAGATGT ATCTATTGAA AAGACATTAG TGGCATCAAA 1200  
 ACGTTGGATT AGATATATCC ATGAAGTTGA AGAGCCAGAT TTTTGTATTG TAATTTATCA 1260  
 30 TGGTGGATTG AATAAAATTA GTAATAGTAC GAAAAATAAA AAGGCAAGTT CGAATGAAGC 1320  
 TGAAAAATTA ATGGAAGAAC TCGGTGTTAT AGATTTAATG ATTACAGCTC ATCAGCATCA 1380  
 AACAAATAGTA GGTCAAGATC ATGAAACGTA TTATGTTTCA GCTGGTCAAG ATGCCAAAGA 1440  
 35 GCTTGATCAT CTTCGATTA ATTTTAAAAA GAGAACAACA ACTTATGATG TTGAAAGCAT 1500  
 TGATTCTAAA GTGATTGACT TAAATGAGTA TGAAGAGGAT CAAGAATTAT TAGATTTAAC 1560  
 ATTCTATGAT AGAAAAGCAG TGGCTTATTG GTCACAGGAA ATCATAAGTG ATAAAGGTTT 1620  
 40 GATGTTATCA GTAAATGGGT TACAAGATTT AGTCTGTCAA ACACATCCAT TTTCGCAATT 1680  
 ATTACATGAT GCAATTCACC TTGCATTTGA TAATGATATA ACATGTGTCC ACGTGCCTAT 1740  
 45 GaACGGAGAG AAGGGGTTGA GTGGACAGAT TCGAAATGAA GrTTTGTATC aTGCATACCC 1800  
 ATaTCCAGAT aAGCCatGGG tATGACAATT aGTGGTCAAA ATaTCAAGGT ATnTTGGGGT 1860  
 ATAGTTATTC ACCATTAGG 1879

50 (2) INFORMATION FOR SEQ ID NO: 421:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2710 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 421:

	AATCCCCCT TTTCCCCAT CCATTTnCC ATCCACTGGT CCTTACGGGA CCATATTATT	60
	TnAAATTGGA nAAAATTATt TTAAAGaTTA TwActACTCT TtAATCATTt TrGTGAATTA	120
10	AAAAAAGTAG TGCAAAAAGC AAAATATACT TTATACACTA CAAATCATTt ATTTATAATA	180
	AAGTTTCACC AAAAAATGTT CCAACTAATG AAACCGCTTG TTCAGCAGTA TGATTATTAC	240
15	TGTCAATCAA TGGATTTACT TCAACTAAAT CCATTGAGGA AATTAAATGT GATTGATGCA	300
	GTAATTCCAA TGCAAAATGG CTTTCTCTAT AACTAAGACC ACCCAAACT CTAGTACCAG	360
	TGCCTGGCGT TTCAAGCGGA TCTAAAGCAT CAACATCTAA AGATAAATGA ACGCCATCAA	420
20	CATTGCGTGA CTTCAAATAT TCTATTGTAT TTTCAATTAC TTCCTTTATC CCCAATTAT	480
	CAATATCTGA CATAGTAAAT GTTTTAATAT TATGATCTTT GATAAATTGT CTTTCACCTT	540
	TATCTAAATC TCTCATACCA ATTAGTACGA TGnTTTCTGG CTGATTACA TTACnATTTA	600
25	ATTCTAAAAG TTCTTTGGGA CCTTCGCCTG TCAAAATCCT TAGAGGCATA CCATGAATAT	660
	TTCCACTTGG TGACTCTTCA GGTATATTTA AATCACCATG TGCATCATAC CAAATAACAC	720
	CTAAATTATT ATAATGTTTA CTTATTGCTG ATACTGAACC TACCGCAATA GAATGATCTC	780
30	CACCAAGAAC TAGAGGAAAT CTGTTATTTT CAATTGATGC TGAAACCTCT TTATTTAATT	840
	TTTGATTAAC ATCTATAATT TCATCATAAT TTCTTAATCC TTTTGTTCa CTATGAAATT	900
	TTTCAATGTT CACAGCAGGt ACCTTAATAT CCCCTTGTC ATATACATCA AGGTCTAATT	960
35	GCTTTAATCT TGAAATTAAT CCAGCATATC TAATTGCTGT TGGTCCTAAA TCAACACCTA	1020
	ATTTTCTTTG TCCAAATGTT GATGGTGCAC CTATAATATC AATTGCTTTT GTCTTTGTCA	1080
40	TAAGCGTCCC CCTTTGCTCT ATGTAATTAA AGAATAATGT ATGCGCTTAC CATTATCAAG	1140
	CAATAGCTAC ACATATAATC TGTTTATCTT ATTACTTCAT AAAAAAGGT TCTTCATCTT	1200
	TTATGGTGGG AAGGTAAAC TTCCTGCTTT TTTAATACA CAAAAAGCGC AATTGCCTCT	1260
45	ATAATTTAAA GTGACCAAAC CCAAACTAAA GGAGACAAGT GCGCCTATGT GTAATGATAC	1320
	CTTAGAATTA CTAAGAATAA AAGATGAAAA TATAAAATAT ATAAACCAAG AAATTGACGT	1380
	CATTATCAAA GGAAAAAAG CAACAGTGGT TAATGCTGTA CTAACGTATA AGCCTTCGGC	1440
50	CTGTTATTGT TGTGGAGTTA AAAATGAAGG ACAAATTCAT AAACATGGTA AGCGTGT TTC	1500
	TCGTATTACT TTACTTAAAA CTCAAGGGTA TAACACATAC CTCAACTTAG CTAAACAACG	1560

55



EP 0 786 519 A2

	GTGTTTTATC TCAAGATGTG TTA	CTCAAAA AGTTATAGAA GAAGCTACTA AAGTTAAAAC	1680
	AGAGATTGAT ACTGCAGAAG ATA	ACTGTAT CTCTCCATCT ACTGTAAGTC GTATTAGAAC	1740
5	TAAAGCGGCT AATTCATTAC GA	ATTAAACC CTTTAATTGT TTGCCAGAAC ACATCGCTAT	1800
	GGATGAATTT AAAAGCGTTA AAA	ATGTAAC TGGATCAATG AGTTTCATTT TTATAGATAA	1860
	TGATACTCAT GATGTTATAG AT	ATTTTAGA AAATAGAACT ACAAGATTCT TGCCTGCCTA	1920
10	TTTCGAGCGA TTCGATTAA AAA	ATCGACA ACAAGTTAAG ACGGTTACTA TTGACATGTA	1980
	TGAACCTTAT GTCCGATTAT TTC	CGGACCT ATTTCTAAT GCAGCTATTA TTTTGGACAG	2040
	ATTCCATATC GTTCAACATT TAA	ATAGAGA ACTTAATAAG TATCGTGTAC AAGTTATGAA	2100
15	TGAATACCGT AATAAAAAAG G	ACCTGATTA TACAATTTT AAGAATAACT GGAAAGTCCT	2160
	ATTGATGGAT ACTAGTAAAA C	CATATTTAG TAAATACAGA TGAATAAAT CTTTAAAGGC	2220
	TTATAACGC TCATCTGACA TT	GTAGAATT CATGCTTTCA AAAGACGATA TACTACGACA	2280
20	CTCCTACGAA CTTGTCCAAG G	ATTACGAAA AGACCTAAGG TTATGTAATT GGCCTAAATT	2340
	TATTAATCGT TTAATTCAG T	AGTAAAA GTCTGTGAGT AAGGGTGTAT GGAAAGTGGT	2400
	TAAATATTAT AGAAAACATC A	AGGATGTT AAGAAATACA ATTTATTACC CAGCATTTAA	2460
25	TAATGGTGCT ATAGAAGGAA T	AATAATA GATAAAATTA ATCAAGTGAA TTTCTTTTGG	2520
	TTACAGAAAT TTCAACAACT T	AAAGCACG TATAATGATG aTTTTCAGCT TGTACAAAGG	2580
	AGaAAAAAAG AAGACAACCA A	GCCCAATAA TGGACTGGCC GCCTAATAwA nGGGskCTAA	2640
30	AAGTTgTATT TTTAAAAATA G	TCCTTTAA ATTATATACC CnCCACATTT GGGGGAGGAC	2700
	CTAAAAAAGC		2710

35 (2) INFORMATION FOR SEQ ID NO: 422:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1027 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 422:

45	CATTTTAATT GtTaAAATTC CAAAAA	tGT ArGTGGATTw AAAGrAAACC CtGTGT	TTTTT	60
	AAAAGGtACC ATTaAAATAg TTCCG	ATTGT ACCATCCCAC GTGAAAtTTT TagT	AtTGCC	120
	GGTGaGAGAA AATGCCAATG CAATC	ATCGC AGTTCATAAT CATCCATCCG GTG	ATGAAC	180
50	GCCCTCACAA GAAGATATCA TAACA	CAAT GAGGTTGAAG GAGTGTGGTT TG	ATTTTAGG	240

55

EP 0 786 519 A2

GGGTTACTTT GATGAAAATG ATTGAAGTTG ATAGAATTAA TGACGTATCT TGTGTATAAT 360  
 ACCTACGAAG TACTTTCATT GGAGGAAAAA TAGTGACTCT ATTTATTATT ATCGGGGTTC 420  
 5 TCGTGCCAAT GGTTTTATACC ATGCAGTTAA ATATTAAAAA TGAACCTGTa ACAAAGCGCA 480  
 ATCTTTTAAT AACATTAGCT TTATCTACGT TAGGTATTTT AGTAACCGCG TTAGCAGGTG 540  
 TAATCGTTAC GAAACAAGCT TTTCTTTTAT TAAGTGTAGC AATTGGCTCA ATTTTTACTG 600  
 10 GAATCGTTTG GGGCCTTTTA CTAAGTGGA GcTACGCGCT GATACGATTT TTATCTAACC 660  
 CATTGGGCG TAAGTAATAG TAATCTGTTT ATCAAGTAGT ATCCGTGCTT GAAAACAAAC 720  
 TAAAACTCCT AATGTGGAAC TAGTTAAAAA ATTCTAACT AGTACATTAG GAGTTTTGTT 780  
 15 ATGCAGAATA AATTTAATG TTAAATTGAA AGTGCGGTAA AAATCCACTA TTTATTTGAA 840  
 AAAAATCGAG CAACCAAATT AAACCTTGAT ATCCTAAGTA AATACATAAT AAAACAAGTC 900  
 CGACATAAAT TAAAAATCGC AAAATAGAAA GTCCAACCTCT AAAAAGGAGG ATGACTAATA 960  
 20 GTGCTATTAA AATAATTGTt AGTATACTCA ATGCTACAAA CCTCCTArTA CGCTTTTAAA 1020  
 TCCATAA 1027

(2) INFORMATION FOR SEQ ID NO: 423:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 569 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 423:

CGGCACATTT TTAATTTATA CATATTTTAA AACTAAGTAA CAGTTTGAAG AAATCGTAGT 60  
 TCAATAATGT TAATTGTGAA AATGTATATA AACATAAAAA AATCATGTAT AATATATGTT 120  
 GTTAATTAAA CAGTTCGAAA GCGAGATGAC ATTATGGGAC GTAAATGGAA TAACATTAAA 180  
 40 GAAAAAAAGG CCCAAAAAGA TAAAAACACA AGTAGAATAT ATGCGAAATT TGGTAAGGAG 240  
 ATTTATGTTG CAGCAAAATC TGGTGAACCC AATCCAGAAT CTAACCAAGC TTTAAGGTTG 300  
 GTGCTTGAAC GCGCTAAGAC ATATTCAGTG CCGAATCATA TTATTGAAAA AGCAATAGAT 360  
 45 AAAGCTAAGG GTGCTGGAGA CGAAAACCTT GATCACCTAA GATATGAAGG ATTTGGCCCA 420  
 AGCGGATCAA TGCTAATTGT TGATGCGTTA ACAAATAATG TAAATCGTAC TGCCTCTGAT 480  
 GTGCGAGCTG CTTTGGTAA AAACGGCGGT AATATGGGTG TATCTGGATC AGTTGCTTAT 540  
 50 ATGTTGTCAT GTGGCCACAT TTGGTATTG 569

55

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3264 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 424:

10	TGGACCTATA AAATGGATGC ATCCAAATAT GGATGnGGGG TanGGCGGGa AAATAGGaGG	60
	TTATATAGTA AGTCATTTAA GTCGAATAGC CATCCTTTTT AAAATGAAAA AAATAGAAAG	120
	CTCAATAGTT TGTAAAGCC TTAAATAGC GTCGTTACGC AATTTTAGAA TGCTAAAAAT	180
15	TGTCACAAAT TTCAAATTAC GTGCTAAAGC TTGAGATATC AATATTTATT GGCGATAGAG	240
	TGTAATTGA CTCCGCTTAC ATCTCAGTTT TATGTTTGA ATTGCTAGC ATAATATTAT	300
	AATAAAATAA AATTGTTAAT CTTAATTTT AGTATAGATA TTTTACGTG TAGTCACGTG	360
20	TAAAATAAAT TCAATTAGGT TAGGAGACAT AATTATGAAA ACATTTGGTA AAAAGGTTGT	420
	ATTAATCGGA GATGGATCTG TAGGATCAAG CTATGCCTTT GCAATGGTTA CGCAAGGTGT	480
	TGCTGATGAA TTTGTAATTA TTGACATTGC AAAAGACAAA GTAAAAGCAG ATGTTCAAGA	540
25	TTTAAACCAT GGTACAGTCC ACAGTCCTTC ACCAGTTGAT GTGAAAGCAG GTGAATACGA	600
	aGaCTGTAAA GATGCAGATT TAGTTGTTAT TACAGCTGGT GCACCTCAAA AGCCAGGTGA	660
	AACACGTTTA CAATTAGTTG AAAAAAATAC TAAGATTATG AAGAGCATCG TTAAGAGTGT	720
30	TATGGATAGT GGCTTTGATG GATATTTCTT AATCGCGGCA AACCCTGTAG ACATTTTAAC	780
	AAGATTTGTA AAAGAATATA CTGGATTACC AGCAGAGCGT GTTATCGGTT CAGGTACTGT	840
35	ATTGGACAGT GCACGTTTAC AATATTTAAT TAGCCAAGAA CTTGGTGTTG CACCTTCAAG	900
	TGTTGACGCT AGTATTATTG GCGAGCATGG TGATACTGAA CTTGCAGTTT GGTCAACAAGC	960
	AAATGTAGCA GGTATTTGAG TATATGACAC ATTAAGAGAA CAACTGGTA GCGAAGcTAA	1020
40	AGCGGAAGAA ATTTATGTGA ATACACGTGA CGCTGCTTAT GAAATTATCC AAGCTAAAGG	1080
	GTCAACATAC TATGGTATTG CATTAGCATT GATGCGCATT TCAAAGCCA TTTTAAATAA	1140
	TGAAAATAAT GTCTTAAATG TTTCTATACA ATTAGATGGT CAATATGGTG GTCACAAAGG	1200
45	CGTTTACCTA GGTGTACCAA CATTAGTTAA CCAACATGGC GCAGTTAAAA TTTATGAAAT	1260
	GCCATTAAGT GCCGAAGAAC AAGCGTTGTT CGATAAATCT GTTAAAACAT TAGAAGATAC	1320
	ATTTGATTCA ATTAAATATT TATTAGAAGA CTAAGCCTAT TTTAAGTATT AATTAGAAAT	1380
50	ATATTAATGG TAAGAGGATC TATTAGTGTT GCAGTAACAC GTGGCACTGA TAGGTCCATT	1440

55

	TATGACTCGT	AAAAATTTAG	TGACATGAAT	AAAAATGTTG	AGGCGTCATT	GAGTATAAAG	1560
	GAAAGTAGTT	CTGCATTAAT	CACGAaGTAG	AGCATGACmA	CGAGGAATAA	CTATAGGGaG	1620
5	ATGGTTTTGG	AATGACGAtG	TCTTGTATCA	ACATGGTACA	TTAGGTACGT	TAATGGCTGG	1680
	CTTACTAGAA	GGCACAGCTA	CAATTAATGA	ATTGTTAGAA	CATGGGAATT	TAGGGATTGC	1740
	AACGTTAACA	GGGTCTGATG	GCGAAcAATA	TTTTTAGACG	GAAAGGCATA	TCATGCTAAC	1800
10	GAGCATAAAG	AATTTATAGA	ATTAAAAGGC	GATGAGAAAG	TACCGTATGC	ATCGATTACT	1860
	AATTTTAAAG	CGAGTAAGAC	ATTTCCATTG	CAACAATTAT	CACAAGATGA	TGTATTTGCA	1920
	CAAATTAAAA	ATGAAATGTT	AAGTGAGAAT	TTATTTTCGG	CTGTTAAAAAT	TTATGGCACA	1980
15	TTTAAACATA	TGCATGTACG	AATGATGCCT	GCTCAGCAAC	CGCCATATAC	ACGTTTGATT	2040
	GATTCAGCAC	GCAGACAACC	TGAGGAAAAA	AGACAAGATA	TTCGTGGTGC	CATTGTTGGA	2100
20	TTTTTTACAC	CAGAATTATT	TCATGGCGTA	GGGTCTGCTG	GTTTTCATAT	ACATTTTGCG	2160
	GATGATGAAA	GAGCTTATGG	TGGACATGTT	CTTGACTTTG	AAGTGGATGA	CGTTGTCGTT	2220
	GAGATACAAA	ACTTTGAAAC	ATTCCAACAA	CATTTCCCGG	TAAATAACGA	GACGTTTGTT	2280
25	AAAGCGAAAA	TAGACTATAA	AGATGTGGCA	GAAGAAATTA	GAGAAGCTGA	ATAGTCTAAT	2340
	ATAATTAAAA	GACCTTAGCG	ATATTAGGAA	CAGGTGGTTC	TAAATGCATC	GTTAAGGTCT	2400
	TTTTATATTA	TATATGTGCT	TACATATTTT	GTTGATACGC	GCGTAAAAAG	CTATAAATAG	2460
30	TATCAAGATA	TGGGGTGTCA	AGAATGTGTT	CACGTGCACG	TCGATAAATA	AAGCCTTGAA	2520
	TCGCTTCGAC	TTCAAGAGGT	TGCTGATGCG	CAATGTCGTA	ATACATGCTC	GTTCCCATTT	2580
	CGTCGGGATA	ACCTTGATAT	ATAGTCATAA	TAGTATCGAC	AGTTTGTTCT	GAAAAGTTTA	2640
35	ACCCCTCTGC	CTGTGCAACA	CGACAACCAT	CTAGTAATAG	CTGTCTACAC	AGTGACGTA	2700
	TTTCAGGATT	ATGCATGATT	GCAACGGTCT	GTCTACCGAG	TGCTGTGATA	GAATTAATGC	2760
	CCAAGTTAAC	TAACAATTTA	TACCAAATAG	CTTGTTGAAT	ATTTGCTTCT	AATACAATGT	2820
40	CGATTTGACT	GTCTTGGAAC	AAATCTCTAA	ATTGTCGAGT	TAATGCATTA	TCTTGATAC	2880
	GTAATTGATA	ATCTCTGAAG	TGCGTAACAA	CATCGCCTTT	CTTTTGACCA	CTTATATAGA	2940
	CAACTGCTTG	GCATACGTTT	TTAAATGAAA	TATGTTGAG	TTGACCATAA	CCATTTTGGG	3000
45	CTAAAATGAT	GAGCGTGTCC	TCATGTGCAA	GATGAGTTAA	ATGAGGAATA	ACTGCATCAA	3060
	GTTGATGTGT	TTTGACTGCA	ATAATAATAA	CATCAAATGT	GTTTGTGACA	TCTTCATAAC	3120
	CTTTCACAAC	AATATCTTGT	GCAGGTGCAT	GTGGTACAGT	ATAATATGTT	ATTGTTTTGG	3180
50	CGTGTCTCCC	GATAAGCGTT	GTATGAGGCA	ATGATTGTTG	TAATTCATAG	GCAATAGTTG	3240

55

## (2) INFORMATION FOR SEQ ID NO: 425:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 796 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 425:

```

AGAACGAACC TTTAATACGT ACGTGTGAC CACGAACTGT CATGACAAAT TGATAAAACG      60
AATAGTTTTT CATTAGTTCA TTGTCACATC AATCACTTTT GTkTCACCTT TAATCACAGC      120
ATTTTCATCA TAAATATTAA TTGAAGCTGC TTGATCAGTG TTAGTAATTA TAATTGGTGA      180
AATTACAGAT TTAGCGTTAT TATTAATATA TTCAAGGTTG AATCTTACTA ATGGATCTCC      240
GACGTTAACT TCGTCACCAC TAGACACTAA CACTTCAAAT CCTTCACCGT CTAATTGAAC      300
TGTGTCTAAA CCGATATGAA CTAATAATTC TAATCCGTTA TCTGCTTTTA ACCCAATCGC      360
ATGCTTAGTT GGAAAGACAT TGTCACACG TCCTGCAATT GGAGACACAA CTTCTCCTTC      420
AGTTGGATTA ATACCAAAC CTTCGCCCAT CATTTTTTGT GCGAATACAG GATCTGGAAT      480
ATCTTCAATT TTCACGAATT CTCCAGTTAA TGGTGCATAA ATTGCGATAT CTTTCTGAAC      540
TTCTTTGCCT TTTCCGAATA ATTTTTTAAA CATACTTTCC ACTCCTACTT ATCAAAATGT      600
GATATTAAAT CGCCATAACC CAATTCTTCT AACTTTTCAT ATGGAATAAA TTGAATTGCA      660
GCGGAATTGA TACAGTATCT TAAGCCGCCA CTTTCTTTAG GTCCATCATT AAAGACATGT      720
CCTAAATGAC TATTTGATTC TTCTGAACGC ACTTCAGTTC TCAACATACC AAATGATTG      780
TCGACTAATT CTATAA                                     796

```

## (2) INFORMATION FOR SEQ ID NO: 426:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3393 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 426:

```

ATCCAGCAAA TCTTTCTTAT CACGTTCTGT AATTTTCTTA GTATCCATCT TGATTAGCTT      60
TGATAACTTT TCAGCCGTAT CCAACATTTC CGATTGTGTT GTTTTTCGAC CCCTAGTATA      120
TGTAATAGCC ATTTTAGAAG CATTATCAAC TAAAACTTTC CCATTTCTGT CTAAAATACG      180

```

## EP 0 786 519 A2

	ATGTGAGCCT TGTGCGATTT GTAAATAACC TAAACGTAGT ACTAGTACTG CAAAAATAAA	300
	TACAATCACA CCAAAATATAA AGTTAATTCT CTTGTTAATT GTATTTTGAA CGATTTTCATC	360
5	ATTTGATTTT TCTTTTAGTC TTTTAAACAA AACTACCTAC CTCTATTCAA AGTCTTTCAC	420
	TTTAAATCAT ATATGAATTT AGAAATTATT TCTATCTTTT TGACAAAAAA ATAACGGTCT	480
	CATTGAAGAG ACCGAACAAG TAATCATACT TTATTTTGTT GCATTATATA ATTTCGTCAAC	540
10	TTTTTCCCAG TTAACATACAT TCCAAAATGC GCCAATGTAG TCAGGGCGTT TGTTTTGATA	600
	TTTTAGGTAA TAAGCGTGTT CCCATACGTC TAAACCTAAA ATAGGTGTTT TACCCTCAGT	660
	TAATGGATTA TCTTGGTTTG GTGTAGTCAC AATTTCTAAC TGGCCATTGT TTACGACTAA	720
15	CCAAGCCCAA CCTGAACCAA AGCGTGCAGC TGCTTTGTCA GCAAATTCTT TTTTAAATTC	780
	TTCTAAAGAA CCCCATGTGT CTTTAATTTT TTCTACTACA GTACCTTTTT CTTCTGAGTT	840
	TGGTGAAAGT AACTCCCAGA ATAATGAATG GTTTAAATGT CCACCGCCAT TATTACGTAC	900
20	AGCAGTTTGG ATGTTAGCTG GTACACTGTC TAAATTAGCA ACAATTTCTT CAATAGATTT	960
	AGATTCTAAA TCTGTACCTT CTACTGCAGC ATTTAATTTT GTAACATACG TGTATGATG	1020
25	TCTGTCATGG TGAATTTCCA TAGTTTCTTT GTCAAAATGT GGTCTAATG CATCAAATGC	1080
	GTATGGTAAT TTTGGTAATT CAAAAGCCAT AAATAATCAT CCTCCTAAAA TGTCTGTAAG	1140
	TAAATAATAA CAAGCAGTGA CTGGTTCAAC AAAGAATTTG CTTAAATTCT ACTACTTATT	1200
30	ATTTTCTCTA CTCATTTAAT ATAACTCAAA TCAAAAATAA TTAAACATTT TGTATATAAA	1260
	AAGTTAACAG ATTTGCCATA AAATCATACG AACGGAGTAT GAAATGAACC TTTATCTTCT	1320
	ATAATTTAAA AAATGAGATT TATGCATACA TCGGACCAAA TGTGCATAAA TCTCATTTCT	1380
35	TTATATTAAAT CTGGCAAGA CTCACATACA CCATAAACTT CAAGTTTGTG TTTGTGAATA	1440
	TTAACACCAG GTAGTGATAA TTTTATCTGA TCTATTGGAC AATAATCTAT TACCTTTGTA	1500
	TCTCCACACT TTTCACAGAT AAAATGATGA TGATGATGGT TTGTACAAGC GATTCTAAAC	1560
40	TTCATTTTAC CATCAAGTTC TGTATTTTCA ATAATTCCTA AATCTTTAAA TAAGTGCAGG	1620
	TTTCTATATA TTGTGTCGAA TGAAATTCGA GGATAATTTT CATCCATAAC TTGTTGTATA	1680
	TACTTTGCGT TTATATACTT ATCTTCTTCG AAAAAATAT CTAACATATC TTTACGTTTA	1740
45	TCTGTATATT TTAAACCGTT CTCTTTTAAA ATTTTAATAG CATCATTTGT ATTCATTGAT	1800
	ATTAGCTCCC TTTTAAACT TCATTCGCAT TTTCTGATAA GCCATTGTAA TCATAAGTAA	1860
50	AATAACAAGT AGAACTACAA TTACACCACC CGGAGAAATG TCCATATAGA AAGCTAGGAC	1920
	TAAGCCTAAT ATTACTGATA ATTCACCTAA AAATACACTT AGTAATATCA ATTGCTTAAA	1980

55

TATCCCTACA ACACGCATTG AGGCAGAAAT AACCATCGCT ACAATAACAA TAAATAAAAA 2100  
 TTGAATCCAT TTAGGAATGC CAATGACTTT ACTATATTCC TCATCAAATG ACAATATAAA 2160  
 5 TAATTCTTTA TAAAACAATG TAATAAACAG AACAACTATG ATGGCAATGA CAATAATCGT 2220  
 TGTTAAATCA CTTATATTCA CTGCGCTTAT TGAGCCAAAT AGCAATCCAA CAATTTCTTG 2280  
 ATTGAACCCA TCAGCTAATG AAATGAAGAT TGCACTCAAG GCGATACCAG CACTCATTAT 2340  
 10 AATTGGAATA GCAATTTCTT GGTAAAGCAGT GTATGACGTT CTTAATTTTT CAATTAGAAG 2400  
 CGCACCTACT ATTGCGAATA AGATTCCAAA CCACATTGGA TTAATAAATA CTAGTGTGCG 2460  
 CATAATAGTA AGTAAAAACA TACCGAAAGA TATACCACCT AAAGTTACAT GACTTAGAGC 2520  
 15 ATCAGCTWTA AGTGATAGTC GTCTAACAAc GATAAAAGCA CCGaTTAGAG GCGCAATAAA 2580  
 ACCTAtCAAG ATaCCACTAA TTAAAGAGTA CCTCATAAAA TCAAAATTCA ATAATGCATC 2640  
 TATCAATTGT GACACGCCTT TCCATTTTAA ATAAACTCAA ATCTTTATTA ATTACAACAT 2700  
 20 TCTCGATTAT GCTGATGATC GACAAAACGT ACAGGATGTC CATAAATTTT TGAAATTTCA 2760  
 ACTTCATCAA GTGATTTAAA CTCATCAGTT GTACCATGGA AATGCAAATG CTTATTTAAA 2820  
 CATGCTACTT CAGTAGCAGT ATCTGCTACA ACACCGATAT CATGAGTAAC TAAGATAATG 2880  
 25 GTGATACCTT CTGTTTTTAA TTGATCTAAA GTATTATAAA ATTCACTTAC ATGTTTTGCA 2940  
 TCAATACCAT TCGTTGGTTC ATCAAGTACT AATACTGcAG GTTCTGAAAT CAATGCTCGA 3000  
 GCAATCATTa CACGTTGTTG TTGACCACCT GATAATTCTG CTATATTTTT ATGAATTAAA 3060  
 30 TCACCTATAT TCAGTCTTTC TAGTACTTTA ATCACTTTTT CATTATCTTT GCTATTAAAT 3120  
 GTTTGGAATA GACGTTTTGT CTTTGTTAAT CCGCTTAAAA CAACTTCTTT AACACTTGCT 3180  
 35 GGGAAACCTG AATTAAAGGC ATTTGCTTTT TGTGATACAT AGCTTAATTT AATTGATGTT 3240  
 TcTTATTTTT AAAATCAATA CCTTCAACAA AAATCTCACC ACTTTGTAAA GGTAATAACC 3300  
 CTAGAATCAA CTTCAATAAT GktGATTTAC CAGCACCATT TGGkCCAACA AwTGctAAAA 3360  
 40 ATTCACCTTT ATTTATTTTG AATGnnnTAT ATT 3393

(2) INFORMATION FOR SEQ ID NO: 427:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1123 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 427:

	CCCTGTTTCT TAGCCTTTTC GTATATGCGT TCATAAGAAG AATTCTTTTT GATTTCAGAT	120
	TGTAAAGCAC TGTTTTCACT TGATTGTTTA TCTATTTTAT AATCTAAATC TGCAATCTTT	180
5	CCTCGCGTAT CATACGCATC CATTTTTTAA GATAGCATAT AAATACTTAA CATAGCAATT	240
	ACAGTAATCA AAGTTATGTA TAAAACTTTT TCAAAATTAG TTAATTGTAC AACCACTTTT	300
	CTCGAAACAG TCTTCTTTTC GGGCTTAGTT TGTGGTTGTT GCTTCGGTAT ACTATTATAA	360
10	ACTTGTTTCGT CATATGGTTG GTACACTTTT TCTACAGCCA TTATAAATTG CTCCTTATTT	420
	AAGTATTTCA GCTACACGTA ATTCGCGCT TCGTGCTCTG TTATTGTCAT CTAAATCTTC	480
	TTCTGTAGCG GTAATCGGTT TACGATTAAC ACGCTTTAAC TTAGGTGTAT ATGCTTCTGG	540
15	TATAACTGGT AATCCTCTTG GTACCTCTGG ACCTTTTTCa TATTCTTGGA ACACCTGTTT	600
	ACATAAACGA TCTTCTAAAG AATGGAAAGT GATTACCGAA ATCCTGCCAT CTACTTTCAC	660
	TAATTCATC GCTTGTTCTA TTGAATCTTC AAAAGCTGAC AATTCATCGT TTAATGCAAT	720
20	TCGTAGTGCT TGAAATACTC GTTTTGCAGG ATGTCCGCCT TTTCTTCTTG CTTTTCAGG	780
	AATACCTTCT TTTATAATGT CAACTAATTC TAATGTTGTT GTTATTGGTT GTTGTTGCG	840
	ATGTGCTTCG ATTCTTCGAG CTATCTGTTT TGAAAATTTC TCCTCGCCAT AGCGATAAAA	900
25	AATCTTCACT AACGCTTCAT ATGACCAATT GTTAACAATT TCATATGCTG TTAGTTCTTG	960
	TGTTTGGTCC ATACGCATGT CTAATGTTGC GTCATGGTGA TAACTGAATC CTCGTTCTGG	1020
30	AATGTCGAGT TGTGGGCTTG AAACACCCAA GTCGTAATAA ATTCCATCTA CTTTTTCAAT	1080
	GTTTAAGTCT TTTAATATTT GAGTTAATTC ACGGAAGTTG CTA	1123

(2) INFORMATION FOR SEQ ID NO: 428:

35	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 838 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 428:

	TATATCCTCC AnACCATAAA AATGGAATTA TAGTGAAGAC CGCTATGATC CAGGATGACC	60
45	ACGTATTATA GAATCTTTCg ATAGAGAAAC GAATCaACGT ATCCgACATC aCGAAATGAA	120
	ATTAGAAGAT TATAAAGATG AGTTAAGAAG AGAATATCTA AAACAATCTG ACAGAATTGA	180
	AGGAGATGAA TAAGCGTGGG ACTTGATTTT AGTGGTTTAC CAGATTTAGC AGTATTGGAA	240
50	CAAATGAAGG AAAAAGAACA GATTAGTGAG GTTATTGCGC CTGAACATGT TCGTATGCAT	300

55



CATTTCAAAA AATTTGAAGA TGATTTTAAA AATGCGGCAC AAGGGGCTTG GGTGAAAAAT 420  
 GCCACAGACG AATTAAAAGA TATTAGTAAT GATTTAGAAA AAATTCAAGA TATTAAAGTA 480  
 5 TAAAAAGGTA TTAAGAAACA CAATAAGTAT ATAATCCATT TAATAATAAA TGAATTATAT 540  
 AGTTCATAAT TTCGACTATA AGTGGCTATT AGCCATTATT TTTCGGGATC TATGTCAAAT 600  
 CGGACTAATG AATTCAATAA TGGAAGTTAA GCAACCAATC TTTGTTTAAC TTCTTTTTTA 660  
 10 TTTTTTGGAA AATAAAGTTT TGAACATAAT AAAATTTGAT TATGTTTTAA CGAATTTTGA 720  
 TGTTCCTTGA ACTATATATC ATCTAGTCGT CATTTACAAT TGGTAAATAT GACTTCAAAC 780  
 TgkATGAAGG TGATGGCGAT TAAAAGGCTC ATCCGTAGGT TCTAAAGAAC TAGAnTTT 838  
 15

## (2) INFORMATION FOR SEQ ID NO: 429:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1150 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 429:

nGTGTTCAAA TAATAGAATG GCTAATACCC CAACAATAAA TAGGAATAAA TATGTTGATG 60  
 TCGTAATTTT AGTAAATAAT GAAAATCTAA AATTAGTAAC CCTATTTTGA ATATAAGCTT 120  
 30 TAACTTCTAA TAATACTGGG AAGCCGATTG ATCCTAAAAT TATTAAAAAT ATAACATATCG 180  
 CTTGTACAAA ATAATCATGT GCATAAGGAA TTAATGACTT ACCTGTAATA TCTAATCCAC 240  
 35 CATTGGTAGT GGCAGAAATA GAAACAAAAA CACCTTGCAT GATTGCATAT TTAAATCTG 300  
 GATTATCTCG ATAAAAGTAA AATGCTAATA ACATGGCTCC TACGAGTTCG ATTACAAATA 360  
 TTGATTTTAC AATATCAATA ATCAATTTGA CGGTACCACT CATTGTGTTT TTGTTATTAT 420  
 40 CTAACATAAT TAACTGACGT TCTCTAATTC CAATATGTTT ACCTAGTACC ACCCATAACA 480  
 TCGTACCAAT TGCCATGACG CCAATTCCAC CAATATTTAA TATCACGAGG ATAATTAATT 540  
 GTCCAAATGT AGAATAGGTA TCGACAATAC TTATCGGAGA CAATCCAGTA AACTAATTC 600  
 45 CGGATACGGC AACAAATAAT GTGTCAATTG GATTTACTTC TACACCTGGT TTATGaACAT 660  
 ACGGTAAGTT TAATAATAAA AATGCAATGA CAATTGCGAC GATATAGTAC AATACAATAC 720  
 50 CTTGTTGAGG GCTTGATCTT TTAAAAAAT GGCTAAAAAT TGACACGTTG TTCACCTCAA 780  
 CATTATTTTA GTTTAATATT AATTTCTTTC TCTTTACCAT CTCTATAAAT CTTGCTGTGA 840  
 ATTGATTTCA AGTCATCTTT ATGACTAAAT ATAATCTGCC TAAAGCGTAA ATCATCTTCT 900

55

AAACCATTGT TGTCAACTTG ATCTACAACA ACACCGTTCT TAACTTTTCC TGGCAATTTA 1020  
 ACTGCTTGTC TTTCAAAACT ATTTAGACTG ACAATATTCT TCATTTTAAC ACCTACATCG 1080  
 5 GGATAGTCAA TTTTACCTTT tGTTTCTAAA TCTTTTACAA TCTTTtGTAC TTCATTaACA 1140  
 GGtATTGCAA 1150

## (2) INFORMATION FOR SEQ ID NO: 430:

10

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 797 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 430:

20 AAGAATATTT AGGATTTTAT GTTTCGCAAC ACCCAGTAGA TAAAAAGTTT GTTGCCAAAC 60  
 AATATTTAAC GATATTTAAA TTGAGTAACG CGCAGAATTA TAAACCTATA TTAGTACAGT 120  
 TTGATAAAGT TAAACAAATT CGAACTAAAA ATGGTCAAAA TATGGCATTG GTCACATTAA 180  
 25 ATGATGGCAT TGAACTTTA GATGGTGTGA TTTTCCCTAA TCAGTTTAAA AAGTACGAAG 240  
 AGTTGTTATC ACATAATGAC TTGTTTATAG TTAGCGGGAA ATTTGACCAT AGAAAGCAAC 300  
 AACGTCAACT AATTATAAAT GAGATTCAGA CATTAGCCAC TTTTGAAGAA CAAAATTAG 360  
 30 CATTTGCCAA ACAAATTATA ATTAGAAATA AATCACAAAT AGATATGTTT GAAGAGATGA 420  
 TTAAAGCTAC GAAAGAGAAT GCTAATGATG TTGTGTTATC CTTTTATGAT GAAACGATTA 480  
 35 AACAAATGAC TACTTTAGGC TATATTAATC AAAAAGATAG TATGTTTAAT AATTTTATAC 540  
 AATCCTTTAA CCCTAGTGAT ATTAGGCTTA TATAACTTTy ATCAACTAAT AAATTATGAT 600  
 ATAGTAAACT GATGGTTAGA TATTTyTyAA CCATCAGTTT GCGTklATAT TAGTTTTTTA 660  
 40 TGCTTATTAT TTTTATGAGT TTCACTTTAC ATTATTGATT AATCCATAAG AATAATTAGC 720  
 ACAAAAAGCA GTATACATAA ATTGAGTAAA GAATTTTGTC GATATAACCG nGCGGAAAAA 780  
 TAATAACnTT TTGGATn 797

45

## (2) INFORMATION FOR SEQ ID NO: 431:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1466 base pairs  
 50 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

	TTTCCAAGAG AAGTTGCGGA AGTAATTAAT AAAACGCATC ATAATAAATT GGTCAATTCG	60
	ATGATTTTCGk CACAAATCGA TCGGATAGA ATGGATTATT TACAACGTGA TCGTATTTC	120
5	ACAGGTGTAT CATATGGTGC TTTTGATATG GAGCGTATTT TAAGATTAAT GCGACCTTCT	180
	AAAGATGAAG TACTAATCAA AGAAAGTGGT ATGCATGCAG TTGAAAACCTT TATTATGAGT	240
	CGTTATCAAA TGTATTGGCA AATTTACTTC CACCCAGTTA GTCGTGGTGG AGAAGTGCTG	300
10	CTTAATAAATT GtTTGAAACG CGCAAAACAG CTTTATAATG AAGGCTATGA ATTTAAGTTG	360
	CATCCACATG ATTTTATTCC ATTTTTTGAA GAGACAGTTA CGATTGAACA ATATGTTGAA	420
15	CTCGATGAAG CGGTAGTTAC GTATTATTTG GAAAAATGGA CAAAAGAAGA TGATGCTATT	480
	TTAAGTGATT TAGCAAGTCG ATTTATTAAT CGAGACTTAT TTAAATATAT TCCATTTGAT	540
	GGCTCAATTA TTACAATATC AGAACTGCAA GAACTGTTTG AAGCAGGTGG TATTAATCCA	600
20	GATTATTTATT TTGTGAGTGA AGCATTTTCT GATTTGCCAT ATGACTATGA TCGACCGGGG	660
	TCAATCGCA AACCGATTCA TTTATTAAGA CAAGATGGTA CGATTAGAGA AATAAGCAAT	720
	CAATCATTAG TCATTCATAG TATTACAGGC ATTAATCGCC AAGACTATAA ATTATATTAT	780
25	CCTAGAGAAA TGGTTGCAAA GATTAAAGAT AAGACAATTA GAGAAGCTAT TGAAAATTTG	840
	ATTAATGAGC TTAATTAAAC AGGGCTAAAA TTGTTATCGT TAAATATGGA GGTTATATCA	900
	TTGTCTGAGA AAAAAGGCTT TAATTTTAAT ATCATAAAAA ATGACCCTCT AGATGGTCAT	960
30	AAAGGTACAA ATATTGGTTC AATTAGCTTA GACAATATTG CACCAGTTTT TATCGATGTT	1020
	GCTAACAAAG AAGCATTTAT TGATATTGGA GGCATGCATG CTCGTGCCAA AGTTGAAAAA	1080
35	GGTGTGAAAT GGATTACTGA TAAAGCTGCT GTTGAAGGCG ATGAAGCTAA AGaATATTGG	1140
	TTGTGTTGGG TAACAACAGA ACGTAATGAA CAAGGACCAT ATTACGCTGG TTTAACAGCG	1200
	TGCTATTTAT TAGTGAATAA AGCAATTCGT CGTGGTTATA AAAGTATGCC TGAACATGTT	1260
40	AATATGATGG ATAAATCAAT GAAACATCAT ATTATCATAG ATCAAATTGG TGACGAGAAT	1320
	AAAGCTATTT TAAAAGACTT TTTAATGAAC CATGATGAAG GTATGTGGAA GCATTCTTCT	1380
	GATGCTTTAC ATCAAGCATT TAATTAAATA TTAGAACTA AAATTTCCCA ATTAATCTAT	1440
45	AAAGATATGA TTCATTTCTC AATGAC	1466

(2) INFORMATION FOR SEQ ID NO: 432:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2304 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 432:

	TTGTACGAAT ACAAAGTTGC ATTAATATAG AATCATTTTAT GCTTTTTTAAT ATCACAAAAC	60
5	AAATTGATTA AACTAGCTAA AATAGTCAAA ATTGGAATCA AGATTACATA AATATCATAA	120
	CCTCTAGATA TTAGTATACA TAATATAATT AGCGACCCTA TTATAATTGC TAAATTATAA	180
	GCCAAACTAA TTTTATATTT CATTTCAATT CCCACCTTAA TAGCATTTTAT TAAAATTCTA	240
10	TAATGTTTAC ATAGACTTCT TAACAACTT TAACCCAGCT AAGCTAAGTA TTGAAATAAT	300
	TACACCTCCT AATATAAATA TATTACTTAA GCTTAGTAAA GGTATAATTA TACTTATCAG	360
15	TCCTAAAGAC AATGTATCCG CTGcATAATT CGaTgTAGAT GAGATACTAA ATACTTTCCC	420
	CATCAAATGA TTTGGCGTTT TTATTTGAAT GGCAACTGAT CTTGTTAGTC CCTCTATAGA	480
	TTGTCCAAGT CCCAATAATG TTGCACCTAT ATATAATATC GCCACACTTG GAAACACATT	540
20	AATAATCGAT AAGCCAATTC CCCAACTAA AACACCAATA CTAAATTTAA AGATTAATCG	600
	CTTTTCTGAC AGCAAACCCA TAATCAATGA CATTAAATAA GATGCTATAC CTAAACATGA	660
	TGTAGCTAGT CCATATACAC CAACGCCCTC TTTTAATATA TTGGAAATAA ACAATGGTAA	720
25	TACAACACGC CAAAGGCCAG TATTAATCAA TATGCAGGCA AATTGGATGA TTATAATAAA	780
	TGGAATTTCT TTAGATTGTT TCAAGAATTC CCAAGTTTCA GAAAAATCTT CTTTGTAGTG	840
30	TCTATCAATC ATGTTGTTAT TTGTATATTT TAAAAGTGCA TTAAAAATAA ATCCTAAAAA	900
	TAGCAATATA CTACAAATAA AAAAGACGCC AACATTACCA ACTAGTATTA CAATGACACC	960
	AATTAAAGCA GGTAATAATA TATTGAGCC TCTTTGCAAA CTATCGATTA ACGCATTACT	1020
35	TGTTGCTAAA TGCTCCTCAT CAATAATTTT AGGAAGAATT GCCCTAAACG CAGGATCCGT	1080
	ATAGCAGTTA ATAATGGTGA TAGCTGTAGA TATGGTTAGA AGCGTCAGAT AATTTAAATT	1140
	TGATGTTATT GCAAGTAAAG GAATAATTAT TATAATCAAA CTTAGTATAA GATCAGATAG	1200
40	ATAAAGTATT TTCTTTTAT TATGTTTATC AGAATATGCG CCACCGAAAA TACCAAATAT	1260
	AATAAATGGA AGTGTTTGAC TCATAACCAT CATTGATAAT TTTAAAGATG ATTGGTTTGT	1320
45	CAATTCAACA GTAAACCAAA TTATTTGTAA CGAAAACAGC ACAAACAAC TCCGACGTAA	1380
	GAATTACCA ATCAATAAAT ATGTAAAGTT TCTATTTTTC AAAACTTCTA AATACAACAT	1440
	ATTTATCACC TCTCATAAAA ATAATTGAAT GCATCCACCA GCTTTTTTAG ACCTTCTTCT	1500
50	AAACTCTCTT TATCCAAAGC GCAATTAATT CTAATATAAT TTAGTCAGTT AAATATCAAT	1560
	TATTTGAAA TATACATACT ACTTGAAACA CCATACATAA CCCCCAAAT GACTACTCAG	1620
55	AGGTTATATT CTAATAATTA TGATTATATT AAATATGAAA ATATTATCAA AAAAATCAAA	1680

TCATTCTCAT CATATTCTAG GTTGTTTTAA ACAAATAAA TATAGTGAAT GCAAATCAAC 1800  
TATTATTTAA ATTATGAATT ATTTAATTC TTTCTTCTAC GAGCCAATAA CATTAATCCA 1860  
5 GCAATTCCAA TTATACTACT AAAGATCAAA CCTTTTTCG TGCTTTCTAA ACCTGTTTTT 1920  
GGTAATTCTG CTCGTTTTTT CTCTTGATTA GCTACTGATT CTTTAGCAAT TTTAGATTTT 1980  
TTAACTTTAT CATTTTTATC CATTGAATGA ACTGGGCCAT TTGGTTTTGC TCTGTCTTTC 2040  
10 GATAATCCTG GATTGTTAGG ATTTACTGGG CCACTTGGAT GAGTTGGTCT GCTCGGCTTC 2100  
TCTGGGTTTT CAGGTCCTTT TGGATCTTTT GGTTTCTCTC CACCGAACTC TACAATCTTA 2160  
TCTACTGGTT GTTTtGTGAT CTCTTCTGTT GGTTGACCCT CGCCAACCTT TTCACCTGTT 2220  
15 AATGGGktCA CTGTGAwTGG TGktGTGAwT GTCyTACTTC CTGGktGTCC TTCTTGTTC 2280  
ACTCGCTCTT CACCAGGTTG TAAT 2304

20 (2) INFORMATION FOR SEQ ID NO: 433:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4733 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 433:

30 GTCTTAATAT TTTACAAGAT ACAGCCTGAC AATATTTATA TTACAAACCA ATATGGTAAA 60  
GAGGCTGTGT CTCGATTAAA AAGAATTTTA GAGCAAAGTA TACCTAATAA AGAATATATT 120  
35 GATTTATCAG AGATATCATT AATTACATTT TTTAAACCAG AATATAATAA ACAACATGTG 180  
AATGAGGAGT TTAATTCITT ATCTTCTTCG AAAATAAAGA ATTTAGCCAA AGAAAATGAT 240  
GGTATTATTA TTTATTTAAA TTTTAAGGAT GTTAATTATC AACTGaTaaG TGAAGGCAAT 300  
40 ACTTTTTTTT CTGCAAAACC ATATTTGCAT TGTGTGTTCa ATAAAGATAT TTTAAATATG 360  
GATAAAGTTG ATATAGAAAA TTTTTTCAAA AGTATAGAAG TCAAATACAG TAAATAAGAT 420  
CAATTATIGA ATCCACCGAC CAGCAGATTT TTTTAAAAA TTAATACCCC GTTAATACCC 480  
45 TTTGCTTCAA TTTGATGAAA ATCAATGAAA TTCAAAATTG AAGAAATCCT TTAATATCAA 540  
GGTGTACGAC AGTCTATATA ATCATGCGAA ATTCTAAAAT TTTCTGAcGT AAAAAACAA 600  
ATTCTyTAAAG CAGCTCGTCG TTCACCTCAA TTCTCAAAAC GTTAATTGTC GGACGATATA 660  
50 TATACAAAAC ACCTCGATGT TATGTCGAGG TGTTTTTTTG CGTTTGTGTG GGGAATATGG 720  
AATAGAGTAT AAATGAATTA ATGTCTCAGG TATAGAATTA ATTCAACTAT GAATTATTGA 780

55

## EP 0 786 519 A2

	TTAGCTTGCG	GAGACCGTAC	GATAGTGTTT	TAATAATAGA	TATTCAAGGG	AAAGAGCTAT	900
	CGGATGATAA	CCAGATGATA	ACTCAAGAAC	CATGrTTTAA	GAGTGTTATC	AAAATAAGGA	960
5	GTGACTTAAT	TGAAAGATT	C	TATATTTTGG	AAGAAAGCTT	TTATTyCTGT	1020
	GTTGsGATGC	TGGTGTCTT	CT	TTTTTTAGG	TTTTATATTA	AaACAGATAA	1080
	TATTTAATGA	GTATCTTCTT	AATTTGTTTA	GGAAGTCTT	CTATCATTTA	TAAGTATAAA	1140
10	ACCAATCGAT	AAATATAAGT	ATGAAATGTA	TAAGTCACAG	TTAGATCTAA	GTCTTGCTGT	1200
	GGCTTTTTAT	ATCTCCATAA	AACGACAAAT	TCAAGCCCGA	CATAAAACAG	CATTTTCAGC	1260
15	CCACCATAAA	ACGACAATTT	CAGCCCGCCA	TTGACTAAAC	ACCACATCCC	AAAAATATCG	1320
	TAACAATCCT	CTACATCAAT	CAATCCAACA	TCCCTCATAA	TCACAACGCA	CAAAATCTAT	1380
	TCATGCATTT	TTGGAATACT	TAGTATTACA	AATAACGATT	TTTATTCATC	TAACAAAGGT	1440
20	TATATAATGT	ACTGAAGGCA	ATTTTATGT	ATTACAAATC	TAATCGTACA	TGTAAAATTT	1500
	TGATAAACAT	CATTAAATTT	GCGTAACTAT	CATTAGATTA	CAATCACAA	AGTAATTACA	1560
	TGTAATACAC	ATCTATACAT	CACATTTGAA	GGGAAATGaA	TATAAATGAC	TGATAAAAAG	1620
25	TACACTGCAG	CCGATATGGT	TATTGATACT	TTGAAAAATA	ATGGGGTAGA	ATATGTTTTT	1680
	GGTATTCGG	GTGCAAAGAT	TGACTATCTA	TTAATGCTT	TAATTGATGA	TGGTCCTGAA	1740
	CTTATTGTCA	CTCGTCATGA	ACAAAACGCC	GCAATGATGG	CGCAAGGTAT	TGGAAGATTA	1800
30	ACAGGCAAAC	CGGGTGTAGT	ACTTGTTACA	AGTGGTCCTG	GTGTAAGTAA	TTTAACTACT	1860
	GGATTATTAA	CCGCAACATC	TGAAGGGGAT	CCTGTATTAG	CGTTAGGTGG	CCAAGTGAAa	1920
35	CGnAnATGAT	TTATTACGAT	TAACGCATCA	AAGTATTGAT	AATGCTGCGC	TATTAAAATA	1980
	TTCATCAAAA	TATAGTGAAG	AAGTACAAGA	TCCTGAATCA	TTATCAGAAG	TTATGACAAA	2040
	TGCAATTCTGA	ATTGCTACTT	CAGGAAAAAA	TGGCGCAAGT	TTTATTAGTA	TTCCGCAAGA	2100
40	CGTTATTTCT	TCACCAGTTG	AATCTAAAGC	TATATCACTT	TGCCAAAAAA	CAAATTTAGG	2160
	AGTACCGAGT	GAACAAGATA	TTAATGATGT	CATTGAAGCG	ATTAAAAATG	CATCATTTCC	2220
	TGTTTTATTA	GCTGGTATGA	GAAGTTCAAG	TGCAGAAGAA	ACAAATGCCA	TTCGCAAATT	2280
45	AGTTGAGCGC	ACGAATTTAC	CAGTTGTAGA	AACATTCCAA	GGTGCAGGTG	TAATTAGTCG	2340
	TGAATTAGAA	AATCATTTCT	TCGGTCGTGT	GGGCTTATTC	CGCAATCAAG	TTGGTGATGA	2400
50	ATTATTACGT	AAAAGTGATT	TAGTTGTTAC	AATCGGTTAT	GATCCAATTG	AATACGAAGC	2460
	TAGTAACTGG	AATAAAGAAT	TAGAAACACA	AATTATCAAT	ATTGACGAGT	TCAAGCTGAA	2520
	ATTACTAATT	ATATGCAACC	GAAAAAAGAG	TTGATTGGTA	ATATTGCTAA	AACGATTGAA	2580
55							

	CAATTAAGAA CACATATTGA TGAAGAACT GGTATTAAAG CGACGCATGA AGAAGGAATT	2700
	CTACATCCAG TGGAAATTAT TGAATCTATG CAAAAGGTAT TAACTGATGA TACTACTGTA	2760
5	ACAGTTGATG TTGGAAGTCA CTATATTTGG ATGGCACGTA ATTTCAGAAG TTACAATCCA	2820
	AGACATTTAT TATTTAGCAA TGGTATGCAA ACGCTTGGTG TAGCATTACC GTGGGCAATT	2880
	TCAGCTGCAC TTGTGCGCCC TAATACGCAA GTTGTGTCCG TTGCTGGCGA TGGTGGCTTT	2940
10	TTATTTTCAT CACAAGATTT AGAAACGGCC GTACGTAAAA ATTTAAATAT CATCCAGCTT	3000
	ATTTGGAATG ATGGAAAATA TAACATGGTT GAATTCCAAG AAGAAATGAA ATATAACGT	3060
15	TCGTCAGGTG TAGACTTCGG TCCTGTAGAT TTTGTAAAT ATGCAGAATC ATTTGGCGCG	3120
	AAAGGTTTAC GAGTTACTAA TCAAGAAGAA TTAGAAGCGG CAATTAAAGA GGGCTATGAA	3180
	ACAGATGGTC CAGTATTAAT TGATATACCT GTAAATTACA AAGATAATAT CAAACTTTCA	3240
20	ACAAATATGT TACCTGACGT ATTTAACTAA AATAAAGATA AATGTTAAAG AGGAGTGGGA	3300
	GATTTTATGA CTAATGTTTT ATACCAACAT GGTACATTAG GCACATTAAT GGCAGGATTA	3360
	TTAAAAGGAA CTGCATCAAT AAATGAATTA TTGCAACATG GTGACTTAGG TATCGCTACA	3420
25	CTAACAGGTT CAAACGGTGA GGTAATCTTT TTAGATGGAA AAGCTTACCA TGCAAATGAA	3480
	CATAAAGAAT TTGTAGAATT AAAAGGTGAC GAGTTAACAC CATATGCAAC TGTAATAAA	3540
30	TTTGTAGCAG ATACAAGCTA TGAAACGAAA GATAAATCTT CAGAAGCAGT TTTTCAGAA	3600
	ATTAAGGAAA AGATGTTGAG TGAAAATTTA TTTTCAGCAG TAAAAATTC AGGCTTATTT	3660
	AAAAAATGC ATGTACGTAT GATGCCGGCT CAAGAACCAC CTTATACACG TTTAATCGAT	3720
35	TCAGCTAGAA GACAACCTGA ACAAAGTAA ACGTATGTCA AAGGTTCAAGT CGTTGGTTTC	3780
	TTTACACCAG AATTATTCCA TGGTATCGGA TCAGCAGGAT TTCATGTACA CTTTGCGAAT	3840
	GATGATCGTA ACTTTGGTGG ACATGTCTTA GATTTTGAAG TAGAAGATGT TAAAGTAGAA	3900
40	ATCCAAAATA TAGAAACATT TGAACAGCAT TTTCCAATTC AAGATAAAGA TTTCACTAAA	3960
	GCAAATATTG ACTATAAAGA TATTGCAGAC GAAATTAGAG AAGCTGAATA ATGAATCCAG	4020
45	AAATATAATG ACGGTTTATG AAAATTGACT TCATAATGCG CGATTTAGAA ATGATAGTTT	4080
	GTAAATATGA TTAACCATGA CTACAATAGA ACAAATATAT TTATAATTAC GTCTAAGTAA	4140
	TAAAATAAAT CCCTTCACTA TTAGCAGTAG TGAGGGGATT TATTAGGTTT CAGATATTTG	4200
50	AGATTTGCTG TTATGTTTAG ATTATAAATT GTGGTACACA CTCATATAAA ATTTACTATT	4260
	GTATAGGCCA ATCTGTTACT ACGAGAAGCA AACAACAATA ATTTACAAGT TCAATAACTA	4320
55	AAAAGACAAA CGCCAATTTT TCAGCGCTTG CCCTATAAAA CTATTTTCAA ATTATTATTT	4380

TTTCTTTAGA TTCACTACTT TTTTATTACC ATCATTCAAA GTAAGCGTAT AAGTTGCTGT 4500  
 TTGGGCATTA TTAATTTTTT CTGTTGTAAC ACCACGTTGA GAAGCTAATT CATTTTTTTAC 4560  
 5 TTTACTGTCA ATTTCTTGAT AAAGAACATT TTTATTTTCT GGGAAGATAA AGTAAGTTCTG 4620  
 ATGTAATGCA GTAATACCAT CTAAGAAAT TGTGTAAGGA ACAGTGTGAT AACCATCCAC 4680  
 10 AGTCATTTCT TTATAGCCGT TATTACTATC TGCAGATGCT TCGTGACTCG GTA 4733

(2) INFORMATION FOR SEQ ID NO: 434:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1284 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 434:

AATAAATGGT AATTTTATAT CATCAACGGC TACAAATGGG GCAGTTTCAT ATATAAATAA 60  
 ACACATCTAC TCAACTGnTC TTAATTAATT AATAAATGAT TTACCAGATG ACATATAACA 120  
 25 GCCATTATTA GnTTAATTCT AATAGTTTAT TTAAATTTTC TTCGGTTGTC GCCCAACTGG 180  
 TTGCGAATCT AACAAACACGA TGTTGATCAT CGTATTTTTC CCAAACAGCA AATTTAACTT 240  
 30 TTTGTTCTAA CTCTGCTATT TTCTCGTTAC TTAAAATAAA AAATTGTTGA TTGGTTGGAG 300  
 AATCAAAGTA AAGACGATAG CCTTTATTTT TAAACCCGTC TTTCATCTTA TTTGCCATTT 360  
 CGATAGCATG TCTGCTTATA TTAAAATATA AATTGTCCGT AAATAATTCT AAAAATTGTA 420  
 35 TGCCTGTATA CCGTCCTTTT GCTAAAAGGG CACCGTGGAT GCTTGATTCTG AGTGGTAAAT 480  
 TGTTTCGGTT CATTATTTT CGTAAAAACA ATGGCTTCCC CGCATAATGC ACCTATCTkC 540  
 GTACCACCTA TATAAAATAC ATCACAATAT TTAGCGrTGT CTTTAATAGT CATATCTGAT 600  
 40 TGGTCACTCA TCAATCCATA CCCTAATCGT GCACCATCCA TAAATAATGG AAGCTGATAT 660  
 TGCTTACATA CTTTGATAA CTCTTCCAAT TCTGaTTTAG AGTATAATGT GCCATATTCT 720  
 GTAGGATGAG AAATATATAC CATTCCCTGGG AATACCATAT GGTCCTTTTT AAAATCACTT 780  
 45 TTAAATGTCT CCATGTAAGT TTCAACATCT GAAGCACTAA CTTTTCCTTC CTTAGAGGGT 840  
 ATAGTAATTA CTTTATGTCC ACTATATTCA ATTGCACCGC CCTCATGCAC AGCAACATGA 900  
 50 CCAGTGTCTG CTGAAATGAC CCCTTCGTAA CTTTCTAACA TTGAATTAAT AACAACTGA 960  
 TTGGTTTGTG TTCCACCTaC TAAAAACGA ATTGTAGCAT TTGGaCAGTC AATTGTATCT 1020  
 TTAATCTTTT CAATTGCCTG AGCTGTGAAT TGaTCAAAGC CATATCCCGA AGCTTGTACA 1080

55



TCGAATGAAA TCACTACATT TCCCCCTAAA ACTAATATCA ACATTTTAAAT AAGATAAACC 1200  
 AATTTCAAAA CTAGTTCGAT ATTTAAAATG TATTATGGAT GGnTAAAGTT TGTATCGCAT 1260  
 5 TATCGCGAAG TTGnATAAAT ATAT 1284

(2) INFORMATION FOR SEQ ID NO: 435:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1072 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 435:

ACAGCTTTTG GTAaGGAGA AAwTCAaTat GAAACAGTAT AArGCGTATT TaATCGATT 60  
 20 AGATGGCACA ATGTATATGG GAACAGATGA GATTGATGGA GCAAnACAAT TCATCGATTA 120  
 TTtAAATGTA AAAGGCATTc CTCATTTATA CGTAACTAAT AATTCAACAA AAACACCTGA 180  
 GCAAGTAACT GAAAAATTAC GTGAAATGCA CATTGATGCT AAACCAGAAG AGGTTGTAAc 240  
 25 GTCAGCGTTA GCCACTGCTG ATTATATTTc AGAACAATCA CCAGGAGCAT CAGTATATAT 300  
 GTTAGGTGGG AGTGGTTTAA ATACTGCGTT AACCGAAGCG GGACTTGTCA TTAAAAATGA 360  
 CGAGCATGTT GATTATGTAG TTATTGGACT TGACGAACAA GTTACATATG AAAAGCTTGC 420  
 30 GATTGCAACG TTAGGTGTAA GAAATGGtGC AACATTTATT TCTACAAATC CTGATGTATC 480  
 AATTCCTAAA GAGCGTGGTT TATTACCTGG TAATGGTGCT ATTACAAGTG TTGTAAGTGT 540  
 35 ATCGACAGGT GTATCGCCAC AATTTATTGG TAAACCAGAA CCGATTATTA TGGTTAAAGC 600  
 ATTAGAAATT TTAGGATTAG ATAAATCCGA AGTTGCTATG GTAGGCGATT TGTACGATAC 660  
 CGATATTATG TCTGGTATTA ACGTAGGTAT GGATACGATT CATGTACAAA CAGGTGTATC 720  
 40 TACGTTAGAA GATGTGCAAA ATAAAAATGT GCCACCAACG TATTCTTTTA AAGATTTAAA 780  
 TGAAGCAATA GCTGAATTAG AAAAATAGAT ATAGTCATTT TATAAAGTAG GTGAATTGAT 840  
 TTGGTAAAAA TAGTTGTTTC GAGGAAAATT CCAGATAAAT TTTATCAACA ATTAAGTAAA 900  
 45 CTTGGTGACG TTGTTATGTG GCAAAAATCA TTAGTGCCTA TGCCTAAAGA TCAATTTGTG 960  
 ACaGcCtTCG TGACGCAGAT GCTTGTTTTA TTACATTAAG TGAACAGATC GATGCAGAAA 1020  
 50 TTTTAGCGCA ATCACCAAAT TTAAAAGTAA TTGCGAATAT GGCTGTAGGA TA 1072

(2) INFORMATION FOR SEQ ID NO: 436:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3271 base pairs

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 436:

	TAAAACTTA CTTTAACACC ATTCCTTTT AACTTTTTTC GTGTTTCnTT TnCTTAAGTC	60
10	CATCCATATT TTTAATGATG TCATCTGCTG TTTTATCTTT TAAATCTAAC ACTGaGTGAT	120
	AmCGAATTG TAGCACAGGA TCAAATCCTT TATGGAATCC AGTATGTTCA AATCCTAAGT	180
	TACTCATTTT ATCAAAGAAC CAATCATTAC CAGCATTACC TGTAACTCTG CCATCATGAT	240
15	TCAAGTATTG ATATGGTAAA TATGGATCGA TATGTAGGTA TAGACAACGA TGTTTTTTTAA	300
	CATATTTTGA TAATTCATTA AAGAAAAAGT GTACGAGTTC TTGATTTTCA TAATCAATCA	360
	CTGGACCGCG ATTTGAATAA AAATACTTGA ACACTTTCAT AACAGGTACA GCAGTAAGtA	420
20	AGCAAGCTGC AATGACCTCG TTATTATTGT TTTTATTCC CACTAAATGT GTTTCATAAC	480
	CTTCAGCAAG CTTTAACTCA TAGTGGCCAA CAGTTTGCGT GaAATGACTG TATGGCATGC	540
25	TATCTGTAAa GGCACCAAAC TCTTTAGCTG TTAAATTGT AACTTCATT ATCATTACTC	600
	CTATTTGTCT CTCGTTAATT AATTCATTT CCGTATTGc AGTTTTTCTA TTTCCCCTCT	660
	GCAAATGgCA AAAATAATAA ATCTAATCTA AATAAGTATA CAATAGTTAA TGTTAAACT	720
30	AAAACATAAA CGCTTTAATT GCGTATACTT TTATAGTAAT ATTTAGATTT TTGAATACAA	780
	TTTCAAAAAA AGTAATATGA ACGTTTGGGT TTGCTCATAT TACTTTTTTT GAAATTGTAT	840
	TCAATTTTAT AATTCACCGT TTTTCACTTT TTCAAACAGT ATTCGCCTAA TTTTTTTAAA	900
35	TCAAGTAAAC TTAATTATTC AATGTTTGTT GGATAGATTG TAAATATTTA ATGATTTTCCT	960
	CACGCGTGIT AGATTTAAAT CGCTTAACGA TTTGCTACC AATGACAATG CCATCTGCAA	1020
40	CCTCTTTTAT ATCTGCAACA TGTGTGGTG TTCTTATACC AAATCCTGCG ACAACTGGCA	1080
	CATTGGCTAT CGCTTTAATT GACTCAATTT TTCGTTTTAA TTCTGGATGA AACGCACCGT	1140
	TTTGCCCTGT TGTGCGATT C ATCGTCACAG TATAAATAAA GCCTTCCGCA TGGGATACGA	1200
45	TATCTTTTAT ACGTTTGTCA TCAGTAGTCA TCGCAACTAA CGATATGATT TTGACGCCAT	1260
	AGTGAATAAA TTGTTGTTTT AAACGCTGCG ATAATTCATA TGGTAAATCA GGAATAATTA	1320
	AGCCGTAGAC ACCAGTATCT CGACATTTTT CAAAAACGC TTGTTCTCCA TAATGACAAA	1380
50	TAATATTATA ATACGTCATT AATACATAGT TACACTTAAT TTGATCACCA TGTTTTTCTA	1440
	ATTGATTGAA AATATAATCT ATCGTGATGC CTGTTTAAAT CGCTTGTTGA CCTGCTTCCA	1500
55	TGATAACTGG ACCATCAGCA ACCGGATCAG AGAAAGGTAC TCCAATTTCA ATTATATCTG	1560

EP 0 786 519 A2

	GTATAAATAA TTTAGTCATT TGCAAGACCT CGCTCTACCA TATATTGTCT AATTGTTTCC	1680
	ATATCTTTAT CGCCACGTCC AGAAATAGTT ACTACAATAA TATCTTCTTT CGACATCGTA	1740
5	GGCGCTAGTC TTTCAACATA ACTCAGTGCA TGTGCACTTT CAATTGCAGG TATAATACCT	1800
	TCATGTTTTG TAAAGTTGAT TAAAGCATTG ATTGCTTGTC TATCACTAGC ATTTTCAAAA	1860
	GTTACTCTAC CAATGTCGTG GTAATAAGAA TGTTCTGGTC CAATACCAGG ATAATCAAGT	1920
10	CCTGCTGAAA TAGAATGTGC TAGTTGCACT TGCCCATCTT CATCTTGAAT TAAATACATT	1980
	TTAGTACCAT GTAATACGCC AGGTGATCCT TTGCCAATTG CAAGTGCATG TTTATCAGTA	2040
15	TCATCGCCTT GACCTGCGGC TTCAACACCG TATAATGCAA CATCATCTTT AATAAATGGA	2100
	TAAAATGTAC CGATTGCATT TGAGCCACCA CCGATACATG CTACAATTGC ATCCGGAAGT	2160
	CGACCTTCTT TCTTCAATAT CTGTGATTTT ATTTCTTTAC CAATCACACT CTGAAAATCT	2220
20	CTAACAATCG TTGGGAACGG GTCTGGACCT AATGCAGAAC CTAATAAATA ATGTGTATCA	2280
	TCTACATGAC TTACCCAATA TTGCAATGCT TTATTAAGTG CATCCGATAA AGTCCCTTGA	2340
	CCATCTTCAA CTGCCACAAC CTTTGCACCA AGTAATTCCA TTCTAAATAC ATTAAGTTGT	2400
25	TGTCTTTTAA TATCTTCACT TCCCATAAAG ACAACAAGTT CCATATCAAA TAATGCAGCA	2460
	ACCGTAGCAC TAGCTACACC ATGTTGACCC GCACCAGTTT CAGCAACAAG CTTCTTCTTG	2520
	CCCATTCTTT TAGCAAGCAA CGCTTGACCT AACGCATTAT TAATTTTATG GGCGCCTGTA	2580
30	TGATTTAGAT CCTCTCGTTT CAAATATATT TTAGCGCCAC CTAGGCTTTC AGTATATGAT	2640
	GCAGCATATG TAAGTGGTGT CGCGCGTCCT ACATACTCTG ATAAATAGTA TTCCAGTTCT	2700
35	CTTTGAAACT CTGGGTCTGC TTTTGCCTCT TTATAAGCTT TTTTCAACTC AATAATTGCT	2760
	GGCATTAAATG TTTCTGGAAC ATATTGCCCT CCATATTAC CAAAGAAACC TAATTCATCT	2820
	GCTTCTGTTT GTATTTGTTT ATTCATTGTC TCTATCTCCT TTCACAATAT TTACAATTGC	2880
40	TGTCATTTTT TCTATATCTT TTCGCCCATT TACTTCTATA CCTGATGCAA GATCATAACC	2940
	TTGATGTGAT AATTTAAGTT GATTAACTGT TTGAATATTT TCAGAGTTAA TGCCTCCTGC	3000
	TATCAAATAA GGTATGTCTT TTATGTGCTT CAAAATAGTC CAGTCATATG TTTGACCGGT	3060
45	ACCACCATAC GACACTGAGG GTGTGTCGAT AATAAATAAA TCTACGAACC CTTTATATTT	3120
	ATTTATGTTT TGGATTATGT TTTCATCTnG CAGtAAAGCT TTAGTGATTT TAATGCTTGA	3180
50	ATATkTCTTT TTAATTTCTT GTAtAAAAAtC AATAGATTct GtGTAAGTGT ATTGTGkTAA	3240
	wTGaCGwAtg CtTAawACgT GTGCCAATGG T	3271

(2) INFORMATION FOR SEQ ID NO: 437:

55

(A) LENGTH: 1553 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 437:

10	ATAATGAAAT AATACTGTGT TTTATCTGCG AAATGTATCA TTTTCTAATc GTTTCACAGT	60
	AAAATGAAAA GATAAAGTGT GTTTTTACTT GAATTTTGAC TAAAATTACT CTATATTTAT	120
	TAATTGAGCT ATGCTTATTA TTACAATTG ATTACAAATT TTAAATTTGT TAATTGAATG	180
15	ATAATATTAA ATAAAGAAAC TTACACAAGC AAATATGAGT TGTAGCCCAA AATACTTGTT	240
	AAATCAAAGT TGAAAGCTAC AAATAATGAA AATTATAAAC TTGAATCTGA AAGTAATTAC	300
	TATAATTATG ACAATGTAA CTTTTAAACG CACTTATTAA TTAACATCAT AATGTTAATA	360
20	TCTAATTTAT TCAAGTACTT TCGCAAGATT TATTATCTAA ATAACGGGGG AAAGAATCAT	420
	GAGTtCACAA AAAAgAAAAT TAGTCTTTTT GCGTTCTTCT TATTAAACCGT AATAACGATT	480
25	ACCTTGAAGA CGTATTTTTT TTATTATGTT GATTTTTCTT TAGGTGTAA AGGTTTAGTA	540
	CAAACTTAA TATTATTGAT GAATCCTTAT AGTTTAGTAG CACTGGTTTT AAGTGTGTTT	600
	CTATTCTTTA AAGGCAAAAA AGCATTTTGG TTCATGTTCA TAGGCGGCTT CTTATTGACG	660
30	TTCCTATTAT ATGCCAATGT TGTGTACTTT AGATTCTTCT CTGATTTTTT AACGTTTAGT	720
	ACTTTAAACC AAGTAGGTAA CGTAGAATCT ATGGGTGGTG CGGTTAGTGC ATCATTCAAA	780
	TGGTATGACT TTGTTATTTT CATTGATACG TTAGTTTACT TATTCATTTT AATATTTAAA	840
35	ACAAAATGGT TAGACACAAA AGCATTTAGT AAGAAATTTG TTCCTGTCGT AATGGCGGCT	900
	TCAGTAGCAT TATTCTTCTT AACTTAGCT TTTGCTGAAA CTGACAGACC AGAATTATTA	960
	ACACGTACAT TTGACCATAA ATATTTAGTG AAATATTTAG GACCTTATAA CTTTACAGTA	1020
40	TACGaTGGTG TTAAAACTAT CGAAAATAAT CAACAAAAAg CGCTAgCATC TGAAGATGAC	1080
	TTAAcAAAAG TATTAAAtTA TACGAAAcAA CGTCaAACmG AGCCTAACCC rGawTATTAT	1140
45	GGGGTGGcAA rGAAGAAAAA TATTATTArG ATTCATTTAG AAAGTTTCCA AACCTTCTTA	1200
	ATTAATAAAA AGGTTAATGG TAAAGAAgTA ACACCGTTTT TAAACAAATT ATCAAGTGGG	1260
	AAAGAGCAAT TCACATACTT CCCTAACTTT TtCCATCAAA CAGGTCAAGG TAAAACATCT	1320
50	GACTCTGAAT TTACAATGGA TAACAGTTTA TACGGTTTAC CGCAAGgTTC TGCCTTTTCA	1380
	TtaaaaGGAG ATAATACGTA TCAGTCATTA CCAGCAATTT TAGATCAAAA GCAAGGCTAC	1440
	AAATCTGATG TCATGCACGG TGACTATAAA ACATTCTGGA ACAGAGACCA AGTATATAAA	1500

55

## (2) INFORMATION FOR SEQ ID NO: 438:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1419 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 438:

5	ACAAAATCAA TCAGCAAAAT GATTGGATTA AGACGCAAAT TGAGCGTTCA ATGGAAGGCG	60
15	AAACAGTTGG CATTAAATGAT CAAAATATAG AAATATATAG TGAAcGtTgA GATTTATACC	120
	ATACACTCGT ACCTTTAAAT CAAGAATTGC ATAAGTTGCG ACTTAAAACT CAAAACCTAA	180
	CCAATGAAAA TTATAATATT AATGATGTGA AAGTTAAAAA GATTATTGAA GATGAACGTC	240
20	AAAGACTAGC ACGAGAACTT CACGATTCTG TTAGTCAGCA ACTTTTTGCG GcMAGTATGA	300
	TGCTATCTGC TATCAAAGAA ACGAAcTAGA ACCACCATTA GACCAACAAA TTCCTATTTT	360
	AGAGAAAAATG GTTCAAGATT CGCAGTTAGA AATGCGTGCT TTGCTGTTAC ATTTAAGACC	420
25	GCTTG GTTTA AAAGACAAAT CTTTAGGTGA GGGTATTAAA GATTTAGTTA TTGATTTACA	480
	AAAAAAAGTG CCAATGAAAG TTGTGCATGA AATACAAGAT TTAAAGTGC CTAAAGGTAT	540
	TGAAGATCAT TTGTTcAGAA TTACACAGGA AGCAATTTcG AATACATTGC GTCATTCAAA	600
30	CGGTACAAAA GTGACAGTAG AATTGTTTTA TAAAGACGAT TATTTATTGT TGAGAATTCA	660
	AGATAATGGT AAAGGTTTTA ATGTTGATGA AAAATTAGAA CAAAGTTATG GACTTAAAAA	720
35	TATGCGTGAA AGAGCTTTGG AAATTGGTGC AACGTTCCAT ATTGTATCAT TGCCAGATTC	780
	AGGTACACGT ATCGAGGTGA AAGCACCTTT AAATAAGGAG GATTCGTATG ACGATTAAAG	840
	TATTGTTTGT GGATGATCAT GAAATGGTAC GTATAGGAAT TTCAAGTTAT CTATCAACGC	900
40	AAAGTGATAT TGAAGTAGTT GGTGAAGGCG CTCTG GTTAA AGAAGCAATT GCCAAAGCCC	960
	ATGAGTTGAA GCCAGATTTA ATTTAATGG ATTTACTTAT GGATGACATG GATGGTGTAG	1020
	AAGCGACGAC TCAGATTAAA AAAGATTTAC CGCAAATTAA AGTATTAATG TTAAGTAGTT	1080
45	TTATTGAAGA TAAAGAGGTA TATCGTGCAT TAGATGCAGG TGTCGATAGT TACATTTTAA	1140
	AAACAACAAG TGCAAAAGAT ATCGCCGATG CAGTTCGTAA AacTTCTAGA GGAGAATCTG	1200
	TTTTTGAACC GGAAGTTTTA GTGAAAATGC GTAACCGTAT GAAAAAGCGC GCAGAGTTAT	1260
50	ATGAAATGCT TACAGAACGA GAAATGGAAT TATTATTATT GATTGCGAAA GGTACTCTAA	1320
	ATCAAGAAAT TGCTAGTGCA TCGCATATTA CTATTAAAAC GGTTAAGACA CATGTGAGTA	1380

55

## (2) INFORMATION FOR SEQ ID NO: 439:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 608 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 439:

ATGTnGGCGG	ACGTAAAGGA	CATGTTTATA	CTGATGATCG	AGCATTAGAT	ATTGATATCG	60
TTCCGCCTGC	TCAAGCAGAT	GGTAAAGCTA	CTAACCCCGA	ACAATTATTT	GCAGCAGGTT	120
ATGCATCTTG	CTTCAACGGT	GCTTTCGACC	TAATTTTAAA	GCAAAACAAA	GTGCGTGATG	180
CTCATCCAGA	AGTAACACTA	ACAGTGAGAC	TAGAAGATGA	TTCAGACTCA	GAAAGTCCTA	240
AATTAAGTGT	TTCAATTGAT	GCGACAATTA	AAAATGTTAT	ATCTCAAGAA	GAAGCTGAAA	300
AATATTTACA	AATGGCTCAT	GAATTTTGTC	CATATTCAAA	AGCGACTCAA	GGAAATATTA	360
ATGTCGATTT	AAATGTAAAT	GTTGTAGATT	AGCATTAACT	TAAAGAGATT	ATTCAACGTT	420
ATTAATAAAA	TTACATAAAA	ATTCAAATt	stCrAcCAAA	AATTTTGGT	TGGyTATTTT	480
TTCTATTTCG	GATTGAAATT	TCTGGCAATA	TTTAACTGAA	AATGATTGTA	CCTTAGTCAT	540
CATAAATGTG	ACCGGTTCCA	ATACTGGCTT	GACTTCTTCG	CATACCGTCT	ACAAATAAAA	600
GTCCAGTG						608

## (2) INFORMATION FOR SEQ ID NO: 440:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 682 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 440:

CCTTTCAAtT	TCTCCCAAgA	TTTACGCATG	TCTGACCAAA	TGTCAATACA	ATACCTGAAT	60
AATGATTTCT	GGAGTAATTG	CGAAAAATAT	TTAAAAGTTG	CTATAGATCA	ATTTTCAAAT	120
TATAGTATCT	CATCTCAAGT	TTCTAACTAT	CATTTACACAG	TATTACTTGG	AGATCGCCAA	180
AAACCACTTA	TGTATCTAAA	TAAAAATCGC	GGTGGTGATG	GTGGCATAAC	AGGTTATATT	240
ATGATTTATT	TAGTGCCGAG	TACAAGTACA	ATTAATTCTA	TGAAAAGCTT	AATTGCACAT	300
GAAGTAAATC	ATAATATGCG	CTATCAATAT	ATTGATTGGG	ATGGCGGAAG	TTTGATTGAA	360

ATTGGACCGT GGGTAACAAA TACTAATTGA AGTCGTGATA ACGTAAAAAT TAAAAATACT 480  
 ATTTATAATC ATTTACATTT AAAGCATATA TTTGAATCGA tGCCTTATCT CTATGGTGAT 540  
 5 GATATTAATA AACTTCAAGG TAGGCCTATC GTTGGCTTAT CTCATGCTGC CGGGTATGCA 600  
 TGTGGCTATC ACTTGGTAAA ATACTTTTTA CAAAAACAA ACATACCTAT TGAAGTTGCT 660  
 ACAACACTTC CAGCACAAAA AA 682

10

(2) INFORMATION FOR SEQ ID NO: 441:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2574 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 441:

20

ATCTTCTTTT AACATGACAA ATTGCAACAA AATATTAATG ACGCCAAAAG ACATTTTTTC 60  
 ACGTTCAATT AATTCTTCAA CCATTGTCTT TTGCGATATA GTTGGTCTTG ATTCAGACCA 120  
 25 AGAAGCTAAC ATATCAATTG GACTCGTTTG TTCAAGTAAC TCAAACCATT CATCACTTTG 180  
 TGGCTTTGGA TTCACTTCTG AAGATTTGCC CGCCGAAGAT GATGTAGCAG GAGATTTACAC 240  
 CTGTAATTTA GGCATTTGAT TTTCGTGTTT CATTAAAGTAA TACGAGCGTG CTGTTTTACG 300  
 CATTTCTTCA AAGGATAACT GTTGTCCACT TGTAATTGAA TTTAAAATAA CATGCTTCAT 360  
 GCCATCTGCT GTTAAACCAT ATAAAGTCGC GAGTTGTGTA ATTAAACGCT TTGCATCTTT 420  
 35 GGTAACAATG TCTTGACTAA TAAAATGTTT ACCTAACATT TGTCTCAACA TTCAAAGTC 480  
 AAAAGATTCA TTTGATAAAT CGATACCTTG GTACGGTTCA TTAATCGGaa TATCACTTGT 540  
 ATCGATATCT ATTTTTGTAG ACGGCACTTT AAAAACATCA GTAAATTGTC TTGTTACCTG 600  
 40 TTTAAATTCA CTCAAATCAA TTTGTTGATA CTCAAAGTAT TTCTTCAACT CATGAAATCG 660  
 ACGATGCTCG ACTTCACTAT ATAAAAAGAT TGACAACATT GGATCATTA AAAATAAATG 720  
 TGCTGAAGGG GtTGaATTAA TTGGTAAACA AATTGTGTTT CTTGTTCATC ATGTTTGACA 780  
 45 AACGCCTTTA ACAATCCAAT CGCTTCAAGT AAGTCCATTT GTTGTCTAAA CTCTAGTAAA 840  
 TTAATTTTAA GTTCATTCAT AAAAATATAA TGAGAAAGAA TCAATGTTTC ATTATGACTT 900  
 TCTTTAACGA ATTGAGTCAT AAAATGATAT AAACCCACTG CTTGCGTTCC AATTAGCGGT 960  
 50 GTATACAGTC GATTCAATAC CTCTAAATGA TTCGTATTTA AATCAAAGTG TTGCATAACT 1020  
 TTGAATTGAT CCTTTGGTCT TAAGCCGAAT TCGAAGGCTT GTCGTCCCaT TTAAGCnATC 1080

55

AAATTCTTTA TAGACTGATG CAAATCTAAC ATATGAACT TGATCAACAT GCATTAACAA 1200  
 GTTCATAACG TGTTACCTA TATCTCGTGA AGACACTTCC GTATGACCTT CATCTCGTAA 1260  
 5 TTGCCATTCA ACCTTGTTAG TTATGTCTTC AAGTTGTTGA TATCTAACTG GTCGTTTCTC 1320  
 ACAAGAACGC ACAAGTCCAT TAAGTATCTT TTCTCTTGAA AACTGCTCTC TTGTGCCATC 1380  
 TTTTTCACA ACTATAAGCT GACTAACTTC GATATGTTCA AATGTAGTGA AACGTGTTCC 1440  
 10 ACAATTTTCA CATTCTCTTC GTCCTCGAAT GGCATTTAAT TCATCGGCAT GCCTTGAATC 1500  
 TACAACTTTA GATTGTGTAG AATTACATTT CGGGCATTTT ATTACATCAC CCTCTTTATT 1560  
 TTGATTATGC CTAATTATAC TATAAATCTA GAGATGAAAA AAGAATCCCT CAATTTAATT 1620  
 15 CATTTAACCA AATAATGAAA CAATAAAAAA CATTATATCG TTAATTATTA AGTAATTTGC 1680  
 ATGACAATAT TATTGTATTA AAAATAAAAA ACCTAACTCC GAAGTCAGAG TTAGGCTATA 1740  
 20 AATTAATTGT ATTAACTTGC ACTTACAGTT TCTTTTGATG TCAAAAGTGC TCCAATTTGC 1800  
 TCAGCAACAT CTACAACCTT ATTTGAATAA CCCCATTCTAT TATCATACCA AGCAATAACT 1860  
 TTTACTTTAT TCCCTGACAT GACCATTTGT GATTTTGCAT CAATAATAGC TGAATTTGGA 1920  
 25 TTAGTATTAA AATCAACAGA CACTAGTGGT TGATGTTTGA CTTCTATGAT ACCTTCTAAA 1980  
 CCTGCATTTT CAAAAGCTTG GTTTACTTCT TCTGCAGTTA CTTCTTTTTC TAAATCAACA 2040  
 ACTAAATCAA CGAGCGATAC ATTCTTTGTT GGTACACGTA ATGCCATGCC GTGTAATTTA 2100  
 30 CCTTCTAATT CTGGTAATAC TTCTTTTAAA GCTTTCGCCG CACCAGTAGA AGTAGGAATA 2160  
 ATGCTTTCAT TACATGAACG TGCACGTCTT AAATCTTTAT GTGGATTATC AATATTTTTT 2220  
 TGGTCATTTG TAATAGCGTG AACAGTAGTC ATTAAACCAT TAACTATTCC AAAGTGATTA 2280  
 35 TTAAAACTT TTGCAACTGG ACCAATGCAA TTAGTAGTAC ATGAAGCATT ACTAAAAATG 2340  
 TCAAATGCTT CTATATCTAA TTGGTTATCA TTTACGCCTT TAACTACCAT TTGAACATGT 2400  
 40 CCACCTTTTG nAGGACCAGT TAACAAAAsT TTTtTGGCAC CTGCTTTAAT ATGTGCGATG 2460  
 GCTTTATCAC CATGaTTAAA TTTACcAGTT GCATCTATAG CAATATCGAT ATCTAATTCT 2520  
 TTCCATGGCA AGTTTTTCAGG ATTGCGATCA GCAACCAATT TAATTTTATG ATCT 2574

45 (2) INFORMATION FOR SEQ ID NO: 442:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3326 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

55



	CAAGGTACCG	GTTTAGGTTT	GTTCAATTTGT	AAAATGATTA	TCGAAGAGCA	TGGTGGTTCC	60
	ATAGATGTTA	AAAGCGAATT	AGGGAAAGGC	ACAACATTTA	TTATTAAACT	ACCAAAACCA	120
5	GAATAAAATT	GAATATAGTT	ATTCAGAAC	GCATGTTATT	GATTAGAGAC	TCTAATTTAT	180
	AGCATGCGTT	TTTTGATTGA	TGTGGGGAAT	TTTGTATGTG	GATTAGAACT	TAGGGTTTTT	240
	GCGAATATCA	ACTATTAAAT	ATATTACTAA	TTTATACAAA	AATATAAAGT	TTGATAAAGT	300
10	TATTTATTTG	ATTATAAAAA	TAGGGTAAAA	TATAGATATA	TTGTATTAAAT	TAAATTATTC	360
	GAGGTGTCAT	ATGAAAAAAT	TCATTGGATC	AGTTTTAGCT	ACGACATTAA	TTTTAGGGGG	420
15	ATGTTCCACG	ATGGAAAATG	AATCAAAAAA	AGACACAAAA	ACAGAAACAA	AATCTGTACC	480
	AGAAGAAATG	GAAGCTTCAA	AATATGTAGG	CCAAGGCTTC	CAACCGCCTG	CAGAAAAAAA	540
	TGCGATTGAA	TTTGCGAAGA	AGCATCGTAA	AGAATTTGAA	AAAGTAGGTG	AACAATTCTT	600
20	TAAAGATAAC	TTTGGACTAA	AAGTTAAAGC	TACAAATGTT	GTAGGTAAAG	ATGATGGTGT	660
	AGAAGTTTAT	GTGCATTGTG	AAGATCATGG	CATTGTATTT	AATGCAAGTC	TACCTTTGTA	720
	CAAAGATGCC	ATCCATCAAA	AAGGATCAAT	GCGCAGTAAT	GACAATGGTg	ATGATATGAG	780
25	TATGATGGTG	GGTACAGTGC	TGAGTGGCTT	TGAATATCGA	GCGCAAAAAG	AAAAGTATGA	840
	TAACCTATAT	AAATCTTCA	AAGAAAATGA	AAAGAAATAT	CAATATACAG	GCTTTACAAA	900
	AGAGGCAATT	AACAAGACAC	AAAATGTCGG	ATATAAAAAT	GAATATTTTT	ATATTACATA	960
30	CTCTTCTAGA	AGTTTAAAAG	AATATCGAAA	GTATTATGAA	CCACTGATTC	GAAAAAATGA	1020
	TAAAGAATTT	AAAGAAGGAA	TGGAACGAGC	AAGAAAAGAA	GTGAATTACG	CTGCAATAC	1080
35	AGATGCTGTT	GCTACACTTT	TTTCTACTAA	GAAAAACTTT	ACTAAAGACA	ATACAGTAGA	1140
	TGATGTAATC	GAACCTAAGT	ATAAATTATA	TAATTTAAAA	AATAAACCAG	ATAAATCTAC	1200
	AATCACAATA	CAAATAGGGA	AACCCACTAT	TAATACTAAG	AAAGCCTTTT	ATGATGATAA	1260
40	TCGTCCAATA	GAATATGGGG	TGCACAGTAA	AGATGAATAA	AATTAATGAT	AGGGATTTAA	1320
	CAGAATTAAG	TAGTTACTGG	GTTTATCAAA	ATATTGATAT	AAAAAAAGAA	TTTAAAGTTA	1380
45	ATGGAAAAAG	GTTTAAACAA	GTAGACAGTT	ATAATGATGA	TAAGAATAGT	AATTTGAATG	1440
	GTGCTGCTGA	TATTAAAATA	TATGAGTTAT	TAGATGATAA	AAGTAAACCA	ACTGGTCAAC	1500
	AGACAATAAT	TTATCAAGGA	ACATCTAATG	AGGCAATTAA	TCCAAATAAT	CCATTAAAT	1560
50	CATCGGGGTT	TGGAGATGAT	TGGCTCCAAA	ATGCTAAATT	AATGAATAAT	GATAATGAAA	1620
	GCACAGATTA	TTTAAAGCAA	ACAGATCAAT	TATCAAATCA	ATATAAAATA	AAGTTAGAAG	1680
55	ATGCAGATAG	ATTATCAAAT	AGTGATTTTT	TAAAAAATA	TAGAATGGAA	TCAAGTAACT	1740

ATCAAGGAGC GAAACATCCG AATGAAAAAG TTGTTGCTAC TGA CTCAGCA ATGATTCCTT 1860  
 ATGCTGCTTG GCAGAAATTT GCTAGACCAC GCTTTGaTAA TATGATTAGT TTTAATAGTA 1920  
 5 CCAACGATTT ATTAACATGG TTACAAGATC CATTCaTCAA AGATATGCCA GGAAAACGCG 1980  
 TTAACATTAA TGATGGTGTG CCCAGGTTAG ATACTTTAAT AGACAGCCAT GTAGGTTATA 2040  
 AAAGGAAGTT AAATAGAAAA GATAACACAT ACGATACTGT ACCACTAATC AAAATAAAGT 2100  
 10 CGGTAAAAGA TACAGAAATT AAAAATGGAA AAAAAGTAAA AAAGACTATT AACATAACAT 2160  
 TAGATATGGA TGGGCGAATT CCAATAAATG TTTGGACAGG AGATTGATT GCACGTTCTG 2220  
 15 GAAGAGGAAC TTTAATTAAA CTTAATTTAG AAAATCTTGA TCGTTGAGT AAAGTATTAA 2280  
 CTGGTGAAAC AAGTGGTATG TTAGCAGAAT GCGTAATCTT TTAAATGAA AGTTTAAACA 2340  
 TCTCAGAAAA TGAAAATAAA AATTTTGCAG ATAGAAAGAA ACAATTATCA GAAGGATTTA 2400  
 20 AGGATAAGAT TAACTTATTT CAGTTAGAAG AAATGGAAAG AACTTTAATT AGTAAAATAA 2460  
 ACTCACTTGA AGAAGTTGCA GATGAAACAA TAGAAAGTAT TAGTGCTGTT AAACACTTAT 2520  
 TACCTGATTT TGCATTGGAT GCATTAAAAG AAAGAATTAA TGAGTTGTTT AAAGGTATAA 2580  
 25 AATCTTTTAT AGAAAAAGTG TATGATAGTA TAGATAATGA AATTTTAGAA ATTTTCAAAA 2640  
 ATATAGATCA CGACTTCAGA GATGGAGTAT CTGAAGAAAT GATGAAACAT TTGAAAGTAG 2700  
 TGAAACAGAA TATAGACCAA ATAAAAATC AAAATGATAT TTATGGTAGG CAAATTGCAG 2760  
 30 ATATTAGAAG TATTATGAAA CAACAAGATG CAACAATTTT AGATGGAAAT TTTCAAATTA 2820  
 ATTGTAGCGG CGAAAATATG GTACAGGGTC TAGTTATACC TTCTAATTAT TTAGGAAGAA 2880  
 AAATGAAAT ATTAAAAGAC CATATCGATG ATGGTATTAA AAAAATAGCA GACTATGTTT 2940  
 35 AAGGTATATA TGATGAATAT GCATCGAAAA TTGTCGATGT AATAAAATAT TTGATTAATA 3000  
 CAATTCCCAA AATACGTAAG AATTAAAGAC ATGCAATTGA AATGTTAAT GTAAAAAGA 3060  
 40 AAGAATTTTT GTCCCTGATT CCTAATGTAA CTGTGAATTA TATTAAACT AAATTAGAAG 3120  
 AATTAGATAA TACTTTAGGC AAATGGGAGC CTTTCTTAA TGATTTAAAA GCAGTGTAC 3180  
 CAATTTTAGA TAACCATTTA GATGATATTG TTAAGAACAT GAAGCCTTTG ATTGTACAAA 3240  
 45 TGATATwTGA ACCATCACAT TATGaCGATA TGTTTAATTC aAGAAAAGCT TTAACGcCAG 3300  
 TGTCTCAAG CGTTTATAA AGGTTG 3326

(2) INFORMATION FOR SEQ ID NO: 443:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5301 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 443:

5	ACCTATAGCT GGCCCGTCAA GTCCTGAATC ATATGTCGCT ATAAACCCAG TTGCATGCCC	60
	AGCAATCTCG GTTTC AATTT CAAAATCATG CTCTTTCAAA CGATCTATTA AAGTTCGAGA	120
	CGCAAATATT TCTTCATTAC CAAGTTCAGG ACGTTCATGA ATTCTATGAC TGATTTTCGAT	180
10	ATAACTATAT TTATTTGTTT CTATATAATC GAGAATTTGT TGTTTTTCAC TCATTTTTTA	240
	CTATCTCCCT TTTACCCTAC ACTCATTTTA ATCATCCACA ATATTTTGTT CTTTCAAAAT	300
	GAATTATTAC TTATTCTATC GGTTTTATCT CATGATGTCA TCTAGTTTTT CTTTATTAC	360
15	AAAATTTTCT AATAACTAAA aGCCTTTCAT AAATTTATAA AACAGTTTCA AATTGTAAAA	420
	CATAAGCCCT ATTGTTACAA TTTAAGTATT GTTAGAAACA CAACTTACAA CAACTTAAG	480
20	TTTATAATG CAACAAATCA TAAGCGTTTT ATAGTTTGA GTAGTTAATA GGAGGAAATT	540
	CAAAATGACA AAAATGAATG TTGAAAGTTT cAATTTAGAT CaTAcTAAGG TGGTTGCCCC	600
	ATTTATTAGA TTAGCGGGAA CGATGGAAGG ATTAAACGGA GATGTCATT CAAAATACGA	660
25	CATTCGTTTC AAACAACCAA ACAAAGAACA TATGGATATG cCCGGA CTAC ATTCATTAGA	720
	ACATTTAATG GCTGAAAATA TTAGAAATCA TAGTGACAAA GTTGTGATT TAAGTCCTAT	780
	GGGTGCCAA ACTGGTTTCT ATGTATCATT TATTAATCAT GATAATTATG ATGATGTATT	840
30	AAATATTGTT GAAGCAACTT TAAATGATGT GCTAAATGCT ACTGAAGTGC CTGCTTGTA	900
	TGAAGTACAA TGTGGCTGGG CAGCAAGTCA TTCATTAGAA GGTGCTAAAA CTATCGCTCA	960
35	AGCATTTCTA GACAAACGAA ACGAATGGCA TGATGTTTTT GGTACAGGAA AATAAATCTT	1020
	AGTCAATCAA GTTAATCAGA AAaGCAGTCG AACAAATGATT TTACAATCGC CATTGTCCAA	1080
	CTGCTTTTTA TTATGCTTCA AAGTCAAAAA ATCGAACAAA TGAAAAAGTA AAATCTTTAA	1140
40	CATTGTCCG ATTTATTTGA GAACCACTAT AATTTCTTAA TTAGTCCCAT TAACACGAAC	1200
	TGCATAGGTA ACCTTAAATA TAGTTGCCAT GTTGCCAATT GTTTATCTCC TAAAGGTAAC	1260
	TTTTTAACTG CCATATAGAT ATTAGCTGGG AATACAGCTA GCAAGAATAG ATTGATTGTA	1320
45	TTTTTCAAGC ATTGAGATGG TCTTTTAATT AAAAGTGCAA GTCCAAATAA TATCTCAAAG	1380
	ATTCCTGTAA CAAGAACCGC TGTTTTTCGA AGTGGCAAAC ATTTCCGGTAT GATATTTCTA	1440
	AATTGTCGTT CTCGTGTAAA ATGCAATACA CCTATTACAC TAAACCTAT TCCTAATAAA	1500
50	TATCTTAGTA TGTTCAATCA GCTTCAACTC CTATTCTGTA ATGATTTTAT GAATTAATGT	1560
	AGGCGATACA ACATGaTCAG CAATTGTTAT GCTTGAATCT AATTTTTTAA CAACATCGTC	1620

55

## EP 0 786 519 A2

	AATTTTTTTA TTAAAAACAA TACCAACCGC TAAATCAATA TCATCCTCTT TTGTTAAACG	1740
	TCCCGCTCCT AACATCATCG AAGCGACACC TATATCGTTA GAGACTAATT CAGTCACATA	1800
5	ACCTGATTTT TTAGCTTTAT ATTCAATTTG ATATTGAGCT TGTGGCAAAC GCTCTGGATG	1860
	GTCAATAACA GTTTCGTCGC CACCTTGGTT TTTAATAAAT GTTTTGAATT TTTCTAATGC	1920
	TGCACCTGAA TTAATTGCCT CAATTAGCAA CGCTCTCGCT TCTTCAAGCG TTTCAGCTTT	1980
10	GTTTGCAAGT ACAACCATTT GAGAACCTAA TGTTAATACA AGTTCTGTGA AATCTTTCGG	2040
	ACCTTGTCCT TTCAACGTAT CAATTGCTTC TTGTAAGTCA AGCGCATTGC CAATCGCACG	2100
15	TCCAAGTGGC TGATTCATAT CAGAAATAAT CGCCATCGTA TTACGTCCCA CATTATTACC	2160
	AATACGTACC ATTGCGTGCG CTAATGCTTC AGCATCTTCT AATGTTTTCA TAAATGCACC	2220
	GCTACCAGTT TTTACATCTA ATACAATTGC ATCTGCACCA GCAGCAATCT TTTTACTCAT	2280
20	AATTGAAGAG GCAATTAATG GTATTGAATT GACAGTACCA GTAACATCCC TTAAGGCATA	2340
	TAATTTTTTG TCTGCAGGAG TTAAATTTCC TGATTGTCCT ACAAAGTCCA CTTTATTTTC	2400
	ATTAACCAAT TTCACAAATG TTGCTTCATC TATTTCAACA TGAAAACCAT CAATTGCTTC	2460
25	TAATTTATCA ATCGTACCAC CTGTATGACC TAATCCACGC CCACTCATTT TTGCAACAGG	2520
	AACATCTACA GCTGCTACTA ATGGTGCTAA AACCAATGTA GTTGTATCTC CTACACCACC	2580
	TGTTGAGTGC TTATCTACTT TGACACCTTT AATATCACTC AAATCTATCA TATCACCAGA	2640
30	ATTAACCATA GCCATCGTTA ATGCTGCACG CTCATCATCA TTCATATCTT GGAAATAAAT	2700
	CGCCATTGCT AAAGTTGATG CTTGGTAATC AGGAATATCC CCTTTAACAT AGCCGCCAAT	2760
35	AAAGAAATTA ATTTCTTCCG TTGTTAGTGT ATGACCGTCA CGCTTTTTCT CAATAATGTC	2820
	TATCATCTC ATTTTATCA TCCTTTTCTT AAAAAGCTTA GGACAAAGCA TCTGCGCTTT	2880
	CTCTAGTCCA TTTTAAAAG CACAAGCGAA AATTATTATA GCAAGCTATC GATTTCATTT	2940
40	TTAATATCAC AATTTCAATG CGATGTTATT ATTCTTAAAT AGATTGGTTA TAACGTTAAA	3000
	GTCCCTATTA AATTATCTTA GAATCATCAT GGCATTTATG ATGCTTTAAA GCTGATATCG	3060
	ACATACTTAT ATATGGTTAC GATGTCCCAT GCTTACATAT TTTTATAAAA TTAGTAATCT	3120
45	GAATCTGCTT CTAAACCTTG CATAATTTGa ACGCCTGCGC TCGCACCAAT ACGTGTGCGA	3180
	CCTGCTTCAA CCATTTTATT GAAATCTTCT AAATTACGTA CGCCACCTGA TGCTTTTACT	3240
50	TCTATATCAG CACCTACTGT ATCTTTCATT AATTTAACGT CTTCTGCAGT CGCACCAGCA	3300
	CCTGCAAAAC CTGTTGAAGT TTAAACGAAG TCCGCACCAG CCGCTTTTGT TAATTCACCTC	3360
	GCTTTTACAA TTTCGTCATG GTCCAACAAT ACCGTCTCAA TAATCACTTT TACTGTGTGA	3420
55		

EP 0 786 519 A2

	TTTAATGCGC CGATGTTGAT GACCATGTCA ATTTTCATCTG CACCATTTTG AATCGCATCT	3540
	TCTGTTTCAA ATGCTTTCGT CGCAGTTGTC GATGCACCTA ATGGGAATCC TATTACCGTA	3600
5	CAAACATAACA CCTCTGAATC AGCTAGTCGC TCTGCTGCAT ATTTAACATG CGTTGGATTG	3660
	ACACATACAG ATTTAAAATG GTATGCTTTC GCTTCATCGA TGATTTGATC GATTTGCGTA	3720
	CGTGTTGACT CAGGCTTCAA TAAAGTGTGA TCAATCAATT TTGCACTATT CATTTTCTAT	3780
10	CTCCTCCTTT ATGGTTGATT ATAAAAATAC GGTGTGAAAT TAGTTGATTG AGCGTCAGGT	3840
	TCATTTAAAT ATCAGGTTAG ATGTTTCGCTT TTTATGTAAC CGCATACATA TACTATTACA	3900
15	TTAATTCATT TCCCATAAAC AAACAATACA ATTGAACGTG ATATCTTCAT TATGAACGAT	3960
	GACTTGACAA CAAGCTAATC AGGATTATAT TTTTATAATT CTTTAATTCT ATAGTACAAA	4020
	AATTCGCAAA AAAGGGAAAC AAATGTTATC TTAAATTAT TAATGAATAT TAAGGAGAAG	4080
20	ATAACAAATG ACAAAGGTA CACCACATAT TCAACCAAAT GGAGTAAAAA TTGCTAAAAAC	4140
	AGTATTAATG CCTGGCGATC CGCTACGTGC AAAATATATT GCTGATAATT TTTTAGAAAA	4200
	TGTTGAACAA TTTAACGATG TACGTAACAT GTTTGGTTAC ACTGGTACAT ATAAAGGTAA	4260
25	AGAAGTTTCT GTAATGGGTT CTGGTATGGG TATTCCAAGT ATTGGTATTT ACTCATATGA	4320
	GTTATACAAC TTCTTTGATG TAGATACAAT CATTCGTATC GGTCTTGTG GCGCATTACA	4380
	AGAAAATGTT AACTTATACG ATGTTATTAT TGCACAAGCT GCATCAACTA ATTCAAATTA	4440
30	TGTAGATCAA TACAATATTC CAGGTCATTT CGCGCCTATC GCTGACTTCG AGTTAGTAAC	4500
	TAAAGCTAAA AATGTCGCTG ACCAAATCGG TGCTACTACA CACGTAGGTA ACGTATTATC	4560
35	TTCTGATACA TTTTACAATG CCGATCCAAC ATTCAATGAT GCTTGAAAAA AAATGGGTAT	4620
	TTTAGGTATC GAAATGGAAT CAGCTGGTTT ATATTTAAAT GCGATTCATG CTGGTAAAAA	4680
	AGCACTTGGT ATTTTCACAG TAAGTGATCA TATTTTACGT GACGAAGCTA CTACACCTGA	4740
40	AGAACGTCAA AATTCATTTA CACAAATGAT GGAAATCGCT TTAGAAATCG CAGAGTAACT	4800
	TATTTAAATT GACTTTAATT GCTCTTTAAC AATGCGATTA AACTCAAAAA GCCAACACAT	4860
	TCTGGGCGTA TCCCCATTTA TGTGTTGGCT TTTATTTATA TTATTACTTA TCTGTAGATT	4920
45	AGCTTAAGTA AGATTTAAAC ATCCAATTAT GTTTATCTAC TGATGTTTGC ATACCTATAA	4980
	ACATATCTTC TGATACATCA TCGCCAGCAT TACCAGCAAT TTCGATTGCG TTTTCTAATT	5040
50	GTTTTGAGAT ATTTGTGAAG TCTTGTGATA ATTCTTCAAC CATTTGTTCT GCAGAGTAAC	5100
	CTTTCGCAGC TTCTTTAACA ATTGATTGCT CTAAGCATTC AGTTAATGTA CCTACAGGGT	5160
	TTCCTCCTAC CGCTAAAATT CTTTCAGCTA ATTCGTCTAC ATATTGGCTT GCTTCATTAT	5220
55		

AATTGTGTAG CTTTGTGTAA G

5301

(2) INFORMATION FOR SEQ ID NO: 444:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 11466 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 444:

15 GAAGATGAAA GCTATATTGA AGATGATGTT ACTAAAAAGG AAGCTATTTT AAGTATGCAA 60  
 ATTCCTAAAG GTTTCTCTCA AAAATTAAAA GAGAACCGTT TAAAAGAAAC GATACAGTTA 120  
 TATGGTAGAG ATGACTTTAT AGGTGGTATT GCTGTAGAAA TTGTTAGTAG TTCATTATAT 180  
 20 GAGCAGCAAA TTCCTAACAT TATTTATGAA CACCTTGAGG ATATGAAACA GCATCAATCC 240  
 ATCGATGCTA TCAACAAGTC CTATCATAAA CATACACCTG AATCTAAAAT CAAATTTGTG 300  
 TCGCTTACTA AACAAGCACA ACACTCTATT TCAATTAGCT TAATCTTTGC GGTGATTTTA 360  
 25 TTTGTTAGCG CTGTTCAAGT AGTACTTCAT TATCGTTTAA ACCAACAAGC AGCATTGCAA 420  
 CGATTATCAC AATATCATTT AAGCCGTTTC AAATATATA GTACTTATGT AATGACACAT 480  
 ACGATTTTGT TATTGTTGGT ACTATTGGCA GTTAGTCTAT ATTTGTCTCA ACCACTCAGC 540  
 30 TTAATATTTT ACTTAAATC ACTGTTACTT ATATTGATTT ATGAGATAGG TATCGTCTTT 600  
 ATCTTATTC ATATTCAAAC AATAAGTCAT CGATTATCA TGACATTTAT ATATGCACTT 660  
 35 GCTATGGGTA TCGTATACTT GATTATTTTC ATGTAAAGGA GCGTAACTGA TGATAGAAAT 720  
 TAATAACCTT TCAAAGCGTT ACCGTAAACA ACAGATTTTC AATCATTTAA CTATGTCCTT 780  
 TGATAGTAAT CGTTTAACCG TATTACTTGG TGATAATGGT GCTGGAAAAT CAACATTACT 840  
 40 TCGTATGATT GCTGGTATTG AAAAAGCTAA TGATGGAAT ATCAACTATT TCGGCGAAAA 900  
 ATGGAATCAA AGACAATAC aAAATCACAT CGGTTATGTG CCACAAGACA TTGCGTTATT 960  
 TGAACACATG ACAGTGGCTG AAAACATTAA ATTTTTTAAA TCACTTTGTA AAAATCCAAT 1020  
 45 TAACGATACA ACTATCAACG AATATTTACA GCAATTAAAC TTTGATGATA CGTCTGCCAA 1080  
 AGTATCTACA TTGTCCGGTG GGAATAAACG TAAAATTAAT ATATTAGTAG GTTTACTAGG 1140  
 TCAACCTCGA ATTCTCATTT TAGATGAACC GACAGTTGGT ATTGATTAA AATCTAGACA 1200  
 50 TGACATCCAC CAACTACTTA ACATCATGAA ATCTAAATGT TTAATTATAT TAACTACCCA 1260  
 TCATTTAGAT GAAGTTGAAG CACTTGCAGA TGATATCAAG TTAATTGGCC AAGATCCTTT 1320

55

	CCCAAGCTGC GTATGATATC GCAACTTGGG ATTTTCTGTA TTATCTACTT TGCAAGTATG	1440
	ACGTTGGGTC TACTGCATAT TGATTACCGA TGCCACCAGA CATACGTTGG AAGTGACGT	1500
5	GAGGCGCTGT TGAATTACCC GTACTACCTG AATATGCAAT TTGGTCACCA GCTTTGACTT	1560
	TATCACCAGC TGAAACAGTT AAACGATTAT TATGCATATA CCATTGGTAG TTATTACTGT	1620
	TCGCTTCTTT AATCGTTACT TGATTGCCGC CACCATAGTT ACTCCAACCT GCTTGACTA	1680
10	CTGTACCATC AGTTAATGAG TAACTGGTG AATTTTCAGG CATTGCATAG TCGACACCGT	1740
	AATGCGCACC ACCACCGTGA TATTGTCCAT ATGGTTGTAG TTGTTTACGA CTTGTTAACC	1800
15	AGCTTGCGTC TTTCGCATGA CCACTAGcTG TCGCTTTACT TGCTGATCCA CCATTTTGAT	1860
	TAGATGTGCC ATTAGGATAA TTGACCTTTC CATTACCATC ATGGCTGTTA TACGCTTGGT	1920
	TGTTGTTACT ATGTGAATAA TAGCTCGCGT CTGGACCTAC ATTTGATTGA TAACCATATT	1980
20	GATTAATATG CTGTTGGCTT TGA CTGCTGCTG TGTAGTCATT GTTATCTCCT GCTGTTGCTG	2040
	GATTCACATA TGTTTGCCG CTTCCATTGG CATTTCGATT TTTTGATAA CAGTTATAAA	2100
	AATAATGCGT ATGTCCTTGA GCATCTACGA ATGTATAGCT ATATTCTTTA TTATCAAACA	2160
25	TTGCTTGATT CCAGTTACCA TCAGGTGTGT GATGATAATC CCCATTAGAA TCAATTGTAT	2220
	AATAAGTACC ATAAGATACG TCTTGTGATT GTGTTGACAT TTGTGTATGT GCTTGTGGG	2280
30	TGTTTGTCGT TTCTGCTGCA TCTGCTTGAT GCGCCATTGT AAATGTAGCG AAsCCATCGT	2340
	TGCAATCGCT GCTGCTGTTA ATTTTTTCAT GTATAAAACA TCCTCCATTA AAGTTAAAGT	2400
	TAGTTTTCAA TTAAACTGTA CTGCACATAC TAAAAGAATT AGACAACTGA GTAAAGGATT	2460
35	TAATTCTCAT TTCCAAC TAATAATATC CCGAAATGTT TTAATAAAT CATTACATTG	2520
	TCATTACAAA ATAGCCATAC ATTGATATTA AAATGACATC TCTCACTGCA TTCGTTTAAC	2580
	CTTTTATAAA TTTTCAAAAT TAACAACTAA TCGTTTCGTC ATGTTTCGCGA TTCAACGCTA	2640
40	ATGCATGATA GTAATCATCC ATCAAATCAT ATCAACCAAA TTCCATTATC AATCGCTATT	2700
	GATTGTCATT CAACTTTCTA ATAGTGATAT GCTTCTCAGG CTAAAAATC GTCATATCTT	2760
	TTCTATTAAT TAAATCATCT GTGAGCTTTA ATGCTACTAA TTCATTGCTG CCATAATACT	2820
45	TAATATATAA CGTTCCTGTA GTTAAATTTA TTACGGTCTG ATACATCGTA TAGTGATTGT	2880
	CAGCATCATG CGGACGTACA ATTCCAATCG GTATATTTAC CGCATCTAAT AAATAAAATG	2940
50	CATTCATTAA ATCCATTCTT TTATCATTGT TTTGAGCAAT GTTTGCTTTC ATAAATGCCA	3000
	TTCTCACAAA GCGCTCAGTT GAAGTAAATC CACCTGGCAA TCCAAATGTA CCTGCTTCAT	3060
55	TGCCTAAAGG TTCAATCGTT ACACCTTCCA ATAAATTTGC TGTGCTGGA TAAGGAGAAA	3120

	CACCAATAGG ATTATCTTTT ATAACCACTT CACCCTCTTT AAATGAAACT TCGACTGTAT	3240
	GTCCAGTTGC ATCGGAAACA TGATAATGCA ATGGCGGAAC TTCACCGATG TCATTTAAAT	3300
5	ATACAGCTAC AACATGTATT TGGGATGCTT GTTGTTTCAT ATCTTCAATG CTTGTTGTAT	3360
	ATCCCAAAAT CCATGTCACA ATTTCATTTT GCGTAATATT CATCGCGTCC GCTTTGTGTG	3420
	TTGATCCATA TGAACATAA CCTCGGAAAT ATTGTGTTGA AATGGCAACG CcATGTTcCAT	3480
10	TAACACCATC ACCATAAATA AAACCTTCCA TATCTGTTCC TGTGCCAATA AAGCCATATT	3540
	GCGTTTGGCC TGTCGTGCCA GTGCAAGATT TCCAACGATA ATTTCTAGGC GTCACGTGCTG	3600
15	GCGAACCATC TAATGGATAA TCATAATCCA TCGTGCGTCC AAGAAGTACT TGATTATTTA	3660
	AAGTTTGTAT TGTGAATCCT GTGCACATTG TTCTCACTCC TCTGTACCTT CATTTACTTT	3720
	AATCACTTTC AAATAAAGCT GTTTCACTTA AACATACTAT AAAAAATCAA TTATACAAGC	3780
20	AATTAATTGA TATTCATTCT CAATAACTGT GGTATGATAT GTAAGGAAAT CATGACTTAT	3840
	GTGTGAGTGA ACGATCATCT ATACATCCGT TCACTTCATC TCATGACTTT CTATATTTAA	3900
	TTTTTACAAG GAGTGACATC TGTGAATAAC ACACAATCTT CACCACGCAG TAATATTATT	3960
25	ATTGCGATTA TGTGTCTGTC ATTAACATAT TGGTTGTTTG CACAATCATT TATTAATATA	4020
	GGACCTCTCG TTGGTCAAAC ATATCAAACC TCTCCTGCCG TGTAAATTT ATCTATTAGT	4080
	TTAACTTCCT TCGCCACAGG TATCTTCATG GTGGCTGCAG GTGATATTGC TGATAAAATA	4140
30	GGACAACTGA GAATGACATA CATGGGTCTC ATAATCAGTA TGTTTGATC TCTTCTATTA	4200
	ATTATATCGG ACATCACTGC ACTGCTCATC ATCGGTAGAA TTTTACAAGG TCTATCAGCA	4260
35	GCTATCTTGT TACCTTCAAC AGTTGGCGTG TTAAATAATC AATTTAAAGG AGAACATTTA	4320
	AGACGAGCGA TTAGTTATCT AATGATTAGT ACTGTTGGTG GCATCGGCCT AGCTGGTGTT	4380
	ATCGGCGGTT TAATTGCCTC AAATTTCTGGA TGGCAAACGA ATTTTCATCAT TAGTATAGTC	4440
40	ATTGCTTTCA TTGCCATATT GCTTCTAAAA GGCACACCTG AAAAAGTAAG TCAACATAGC	4500
	CACCGTCATC CATTGATTA CAAAGGTATG TCGATTTTCG CTGTTATGAT TGGTAGCTTT	4560
	ACATTATTGT TAACACAAGG ATTCGAACAA GGTGGTTTGA GTACATTTTC AATCATTGT	4620
45	CTGAGCATTT TTATCaTCac TACGTTGATA TTCATCATCA TCGAACGTCG ACATGAAGTA	4680
	CCTTTTATTG ATTTCTCAGT ATTACGCAAC CGTCCGTTCA TTGGTGCAAT TTTAAATAAC	4740
50	TTTGTTTTAA ATAGCGGTCT AGGCGTAACA GTGGTCTTTT TCATATATGC TCAAACACAC	4800
	CTTGGTTTAt CAgcTGCgCA ATCTGGACTT GTTACATTGC CATATGCCAT TGTGGCAGTT	4860
	GCGATGATTC GTTTAGGTGA AAAAGCAACA TTACGTTTCG GTGGCAAATT GATGCTCATC	4920
55		



EP 0 786 519 A2

	TCACAATATG TCATTGCAGT TATCATTGGT TTCGTCATAT GTGCGATAGG TAATGGTTTA	5040
	GTCGCAACAC CTGGACTTAC GATTGCAATT TTCAGTATGC CTAATGAAA AGTTGGTTTA	5100
5	GCTACAGGAT TATATAAAAT GAGTGGTACA TTAGGTGGCT CCTTTGGTAT AGCACTAAGT	5160
	ACTACAGTTT TCAGTATGTT ACAAATAAAC TATGCACCAA GTGTAGCTGC AACCGTAACA	5220
	TTTATAGTCA GCATTGTATT GATGATCCTT GGCTCATTGT CTGCATACAT GATCATTCCA	5280
10	AAAACAGTTA AATCTTAAAT ATAATAGAAG AATTATGTTT CGAAATATCT TTATCACTTT	5340
	AAAATGATAT ACAAGAAATC CAAGAAAAAT AAGCGAACTG AATAAATAAA GATTCAATTA	5400
15	ACGCATCAGT ATTAGGATT CACTCTAAAC GATTAATAGT TTTATAAGAA GGTGTTTGAT	5460
	CTTGAGCTAA CCACATCATT CGAATACTGT CATGAAGTAA TTTCTCTATT CTACGACCAG	5520
	AAAATACAGA TTGAGTATAT GCATATAAGA TGATTTTAA CATCATCTTT GGATGATAGG	5580
20	ATGTTGCGcC ACGATGATGT CTGAATTCAT CGAATTTGCT ATCAGGTATC GTTCAACAA	5640
	TTTCATTAAAC ATGTCGCGAA ATATCATTTT GAGGAATTCT AACAGAAGTT TTTATTGGTA	5700
	GTGTAAGTTG GGCAAAGTGT CTTATTTTTT TAAAGTATTT CAAAGTAAAA TTACATGTTA	5760
25	ATACGTAGTA TTAATGGCGA GACTCCTGAG GGAGCAGTGC CAGTCGAAGA CCGAGGCTGA	5820
	GACGGCACCC TAGGAAAGCG AAGCCATTCA ATACGAAGTA TTGTATAAAT AGAGAACAGC	5880
	AGTAAGATAT TTTCTAATTG AAAATTATCT TACTGCTGTT TTTTAGGGAT TTATGTCCCA	5940
30	GCCTCTTACT CTAATTATAT TCACTATCAA TTAGACAAAA TGGCCATTTT CAAATATCAC	6000
	GCGTTGTTTT TGACCTTGAA TATATTTATT ATAATTCTCT TTTTGAAAAT CAGTTAACAT	6060
35	TAATTTAGAT GTACCGTATT TTAACACTTT TTGCATGTT TCTATTCTCA TTTTCTAAA	6120
	TAACCATCCA TCTTTTAACA CAATACGATT AACAGCATCA TATGATAATT CTAAGTTTTC	6180
	TTTAATTTCA AATGTCTTGA ATGAAATAAT CGTGCACATT AAAACGTAT CACCAAAGTA	6240
40	ATAAACATCT AAATCATCAC GTTTATGTTG TCCAATAAC AAACGACCAT ATTCGAACTC	6300
	TTTTTCTGGA TATTTCAATT CTAAAAAAT AATAATCTCT TCTTCTTTTA ATTTGAATTG	6360
	CATTTAAAAA CATCCTCTCT TAAGTTTTAA CAAGCCTTAA TTAATAATTT TTTCAATCAC	6420
45	ATAGTTCAAT ATACATCATT TCGTTATGTT TTTTAATACT TTGTTCAAAA ACAAATATTT	6480
	TATCTTTTAA AATAATGACT TTTGTATTTT TAATCACAAT AAACATTTTA AAATCTTGT	6540
50	TATCATAATC ATTAAAAGGT ATTAACCTTA ATAATATATT CTCTCGTCTC AACCTTAATC	6600
	GTATACTTCA GACGTCTGTT TGTAGACAAT AAAAGTCATT CACGTCTTCA TATGTCATCA	6660
	AATGTTTATC ATGATATGAT GAATATAATA ATCGGTATA TAACTGTATG ATTAATTACA	6720
55		

## EP 0 786 519 A2

	TGGTGTATCA	GTTACAAC	TGTCACATAT	TTTAAATCAT	AATGATAGTC	GTTTTTCCGC	6840
	AACAACGATA	AAAAACGTAC	ATGCTGTTTC	AGAACGTTTA	GGCTATGCCC	CTAATAAACA	6900
5	TGCAAAACAA	TTGCGCGGCA	GTAAAATTCA	AACTATTGGC	GTCATTTTGC	CTAGCTTAAC	6960
	AAATCCGTTT	TTCTCAGCAC	TGATGCAAAG	TATTCATGAC	CATAAACCAT	CTGATGTTGA	7020
	TTTATGCTTT	TTAACATCTA	CAGCAACTGA	TTTGTATGAC	AATATTAAAC	ATTTAATTGA	7080
10	TCGAGGTATT	GACGGATTAA	TTATCGCACA	ATACATATCA	TCCCCGGACG	CCCTAAATAA	7140
	CTATCTAAAG	AAACATCATG	TACCTTATGT	CGTACTGGAT	CAAAATGACC	ATCAAGGCTA	7200
15	TACAGATTTT	GTTCCGACAA	ATGAATATCA	AGGTGGACAA	CTTGCAGCAC	AACATTTAAT	7260
	AGAACTCGGT	CACAACCATA	TGATAATTGT	TGCACCATAT	GACATGATGG	CGAATATGTC	7320
	GACTCGTGTC	GCTGGATTTG	TCGATACTTT	GCGCGCGAAT	CAATTGCCAG	AACCACAAAT	7380
20	CGTCCATACT	GAATTATCTA	AGCGCGGTGG	GCTAACCATT	GTTGATGACA	TCATGGTTCA	7440
	ATCTGCCACT	GCAATCTTCG	CTATTAACGA	TGAACTCGCT	ATTGGCATT	TACGAGGACT	7500
	AATTGAACAT	GGCATCAGTA	TCCCGAAAGA	TATCTCATT	ATAGGTTATG	ACGACATTGA	7560
25	TTATGCAGCG	TACGTCTCGC	CACCTTTAAC	TACTGTGGCA	CAACCTATAA	CTGATATTGG	7620
	CAAAACATCT	TTAACCTTAT	TACTTCAACG	ATTACAGCAC	TTAGATAAAT	CCATTGATAT	7680
	GATTGAATTA	CCAACGACTT	TAAAAATTCT	TGCAACAAC	GGCTATCATC	TTTCAAAC	7740
30	ACTACGTATC	TTCCGAAATA	TACTCATCAT	TGTTAGGCCC	TTAGCGTTGC	TTTAATGCTG	7800
	AGGGTTTTTA	ATCATAATTA	TTTACTAAG	AAATTAAAT	AATAATGTAT	GAATTTTTAA	7860
35	ATATGATTTA	AACGTTTTCA	GTTTTTATGA	AAACGCATGC	ATTTTACAAA	TAAAAATGGT	7920
	ACGATGGCAC	TGGTAAAACG	TTTTACTAAA	AACAAATCAT	GAGGTGTATA	ACATGAGCAT	7980
	TGTTGCATTA	CTTATCGGGT	TAGGCCCTT	AATTGGCTGG	GGCTTCTTCC	CAACAGTCGC	8040
40	TTCAAAGTTT	GGTGGTAAAC	CTGTACATCA	AATTATCGGT	GCTACTGTAG	GTACGTTAAT	8100
	CTTCGCTATT	TTATTAGCCG	TAGTCACATC	AAGTGGCTTC	CCTACTGGAA	CCAATTTGCT	8160
	ATTCGCCTTA	TTATCAGGTG	CAGGATGGGG	ATTCGGACAA	ATCATTACAT	TTAAAGCGTT	8220
45	CGAATTAGTC	GGCTCATCTC	GTGCCATGCC	AGTCACAACA	GCATTCCAAT	TATTAGGCGC	8280
	ATCTTTATGG	GGTGTCTTTG	CATTAGGAAA	TTGCCAGGC	ATTGGTCATA	AAATCATTGG	8340
50	ATTTACAGCT	TTAGTCGTTA	TTCTAATTGG	AGCGCGTATG	ACAGTTTGGA	GTGAACGCAA	8400
	AGAAGCAAGT	AACGCCAAAA	ATTTACGTCT	TGCAGTGGTA	CTTCTGTTAA	TTGGTGAATT	8460
55	TGGATACTGG	TTATATTCAG	CTGCACCGCA	AGCAACTTCT	ATTGATGGCC	TAAGTGCCTT	8520

	AGCAGAGAAT CCATTCCGTA ATAAAATTAC GTGGTTACAA ATTATTTTCAG GTTTCTTCTT	8640
	TGCATTTGGT GCTTTAACAT ATCTTATTTTC AGCACAACT AATATGAATG GTTTAGCAAC	8700
5	TGGATTTTATT CTTTCTCAAA CATCCGTTGT GCTTGCTACA TTAAGTGGTA TTTATTTCTT	8760
	AAAACAACAT AAAACGTCAA AAGAAATGGT TATTACAATC ATCGGCTTAG TACTCATTTT	8820
10	AGTAGCCGCT TCTGTTACAG TATTTATAAA ATAAGGAGTG TAGATGTCAT GAAAAATCA	8880
	GCTGTTTTAA ATGAACATAT TTCAAAAGCA ATCGCGACAA TTGGTCATTT TGATTTATTA	8940
	ACGATTAATG ACGCTGGCAT GCCAATTCCA AATGATCATC GTCGTATCGA CCTAGCTGTA	9000
15	ACTAAAACT TACCACGCTT TATTGATGTC TTAGCTACAG TGTTAGAAGA AATGGAAATC	9060
	CAAAAAATAT ACTTAGCAGA AGAAATAAAA GAACATAACC CTACACAATT GCAACAAATT	9120
	AAACAATTGA TTTCATCGGA AATCGAAATC ATTTTCATTC CTCACGAAGA AATGAAAAGT	9180
20	AACCTTAGCTC ACCCATTTAA TAAAGGTAAT ATTCGTACTG GTGAAaCAAC GCCCTACTCT	9240
	AATATTGCAT TAGAATCGAA TGTTACTTTT TAAAAGTTAT AACTTGAAAG GAGCGTACAC	9300
	ATGACCAACA AAGTTGTTAT TTAGGTTCA ACGAATGTCTG ATCAATTTTT AACAGTTGAA	9360
25	AGATATGCAC AACCAGGCGA AACATTACAT GTTGAAGAAG CACAAAAGC ATTCGGCGGA	9420
	GGTAAAGGTG CCAACCAGGC TATTGCCACT GCACGCATGC AAGCAGACAC AACATTTATT	9480
30	ACTAAAATTG GCACTGATGG CGTTGCTGAT TTCATCTTAG AAGATTTTAA AGTAGCTCAT	9540
	ATTGATACAT CATATATTAT CAAAACAGCT GAAGCAAAAA CGGGCCAAGC CTTTATCACT	9600
	GTGAATGCAG AAGGACAAAA CACCATCTAT GTTTATGGTG GTGCGAATAT GACGATGACA	9660
35	CCTGAAGATG TTATTAACGC AAAAGACGCT ATAATCAATG CAGACTTTGT CGTGCACAA	9720
	TTAGAAGTAC CCATCCCGGC TATTATATCT GCATTTGAAA TTGCCAAGGC ACATGGTGTG	9780
	ACGACAGTAT TAAATCCTGC ACCAGCGAAA GCATTACCTA ATGAATTATT ATCATTAAATC	9840
40	GATATTATTG TGCCAAACGA AACAGAAGCC GAATTGTTAT CTGGGATTAA AGTAACTAAT	9900
	GAACAATCTA TGAAAGACAA TGCCAATTAC TTTTATCTA TAGGCATTAA GACTGTTTTG	9960
45	ATTACGCTAG GTAAGCAAGG TACATATTTT GCTACTAAAA ATCAAAGCCA ACACATCGAA	10020
	GCTTATAAAG TAAATGCGAT TGATACAACT GCTGCAGGCG ACACATTTAT TGGTGCATTT	10080
	GTCAGTCGCT TAAACAAGTC GCAAGATAAC TTAGCAGATG CTATTGATTT TGGTAATAAA	10140
50	GCGAGCTCAC TCACTGTACA AAAACACGGC GCGCAAGCAT CTATTCCTCT ACTAGAAGAA	10200
	GTAAATCAAG TTAAATGAA TCAAACACAG CTATGATATG AAGGTTTAGC ATATAACATG	10260
	CAACATTCGT ATATCATGGC TGTGCTTTTT TATCTTTATA AAACATCATC TATTAGAAAT	10320
55		

TTTGTAATCT TTTTAAC TTC CAAATTaTCG CATATAAATA TGCTATATTA ATGATAATAA 10440  
 TTATCAATTA AAAGGAGGTT ATGCTATGTC TAAAGAAGCT GGTACATACAT TTTTAGCTAA 10500  
 5 ATTAGGAAAA ACTCGTCTAC GCCCCGGTGG TAAAGAAGCA ACAGATTGGT TAATACAACA 10560  
 AGGGGCATTT TCACAAGATA AACAAAGTGT AGAAGTGGCA TGTAATATGT GCACAACATC 10620  
 TATTTATCTA GCTCATACAT ATGGCTGTCA CATTCAAGGC GTTGATATAA ATAAGAAAGC 10680  
 10 ATTAGAAAAA GCACAGGAAA ACATTTcAGC AGCAGGTCTT GAATCATATA TTCAAGTTCA 10740  
 ACAAGCGAAT GCTGTAAAT TGCCCTTTGA TGACAATCAA TTCGATATCG TTTTAAATGA 10800  
 15 AGCAATGTTA ACAATGTTAC CCATCGCCAT AAAGGAAAAA GCATTACGCG AGTACTACCG 10860  
 AGTCTTAAAG CCTGGGGGTA TCTTGTTAAC ACATGATATT GTCATCGTTA ATGAATCACA 10920  
 TGCCACACAT GTTGTTAAAT CATTATCTGC AGCAATTAAT GTCAATGTCT CACCGCAGAC 10980  
 20 GAAACTTGGC TGGTTAGATT TATATAATCA AGCTGGTTTT AATCATGTGC ATTATCATAC 11040  
 TGGTCCAATG AGTTTAATGA CACCAAAAGG TTTAATTTAT GACGAAGGTA TTGTTGGAAC 11100  
 TATAAGATT ATCAACAATG CTTTGAAAAA AGAAATCGA CCAATGTTTT GTAAAATGTT 11160  
 25 TAAAACGATG ACTAAATTGC GTAAAGATAT GAATTATATT ACTTTTGTCTG CTAAAAAGA 11220  
 GCACTAAATA TAATGCCACT AACTGTACTT TGTATCTATG TTTGACTATC ACTTTAATTT 11280  
 CTTTGTGACA CTAATCATCT ACTTAACAAT ATCGTTATCG TTGATTAGTA AGTCATCAAT 11340  
 30 TTTGGTTAAA GACTTTcATA AACACTCAAA CATTAACT ATACATAGTT AGTGGcATTA 11400  
 TTTTTTyCTn AAAATTTTAA CmTCmCGGGr TtGGGAmCrG AAaTGrtAwT TcGCrmAAaT 11460  
 35 TAwTcT 11466

## (2) INFORMATION FOR SEQ ID NO: 445:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2176 base pairs  
 40 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 445:

TTACATAGTT AACACTAGTT AATCTATTAG TTAACATTAG TTAATAATTA GTTAATTTCC 60  
 50 ATTTGTATTC TCATGTGATA AATTCTAAAA GCATACaATA AATTTAATAT GTAAAAAGAA 120  
 AGGGAATACA CATGAAAAAT AAATATATCT CGAAGTTGCT AGTTGGGGCA GCAACAATTA 180  
 CGTTAGCTAC AATGATTTCA AATGGGGAAG CAAAAGCGAG TGAAAACACG CAACAACTT 240

55

	AAGTATTACA TCTAAAAGGT ATCACAGAAG AACACGTAA CCAATACATC AAAACATTAC	360
	GCGAACACCC AGAACGTGCA CAAGAAGTAT TCTCTGAATC ACTTAAAGAC AGCAAGAnCC	420
5	CAGACCGACG TGTTGCACAA CAAAACGCTT TTTACAATGT TCTTAAAAAT GATAACTTAA	480
	CTGAACAAGA AAAAAATAAT TACATTGCAC AAATTAAAGA AAACCCTGAT AGaAGCCAAC	540
	AAGTTTGGGT AGAATCAGTA CAATCTTCTA AAGCTAAAGA ACGTCAAAAT ATTGAAAATG	600
10	CGGATAAAGC AATTAAAGAT TTCCAAGATA ACAAAGCACC ACACGATAAA TCAGCAGCAT	660
	ATGAAGCTAA CTCAAAATTA CnTAAAGATT TACGTGATAA AAACAACCGC TTTGTAGAAA	720
15	AAGTTTCAAT TGAAAAGCA ATCGTTCGTC ATGATGAGCG TGTGAAATCA GCAAATGATG	780
	CAATCTCAA ATTAATGAA AAAGATTCAA TTGAAACAG ACGTTTAGCA CAACGTGAAG	840
	TTAACAAAGC ACCTATGGAT GTAAAAGAGC ATTTACAGAA ACAATTAGAC GCATTAGTTG	900
20	CTCAAAAGA TGCTGAAAAG AAAGTGGCGC CAAAAGTTGA GGCTCCTCAA ATTCAATCAC	960
	CACAAATTGA AAAACCTAAA GTAGAATCAC CAAAAGTTGA AGTCCCTCAA ATTCAATCAC	1020
	CAAAAGTTGA GGTTCCTCAA TCTAAATTAT TAGGTTACTA CCAATCATTA AAAGATTCAAT	1080
25	TTAACTATGG TTACAAGTAT TTAACAGATA CTTATAAAAG CTATAAAGaa AAATATGATA	1140
	CAGCAAAGTA CTACTATAAT ACGTACTATA AATACCAAGG TGCGATTGAT CAAACAGTAT	1200
	TAACAGTACT AGGTAGTGGT TCTAAATCTT ACATCCAACC ATTGAAAGTT GATGATAAAA	1260
30	ACGGCTACTT AGCTAAATCA TATGCACAAG TAAGAACTA TGTAAGTGA TCAATCAATA	1320
	CTGGTAAAGT ATTATATACT TTCTACCAA ACCCAACATT AGTAAAAACA GCTATTAAAG	1380
35	CTCAAGAAAC TGCATCATCA ATCAAAAATA CATTAGTAA TTTATTATCA TTCTGGAAAT	1440
	AATCAATCAA AAATATCTTC TCTAGTTTTA CATCATTTTT TAAATAATTT TCGTAACAAA	1500
	COGTGATTAA AAAGAACCGT TGATTCTCAA TCGAATCTAC GGTTCTTTTT TCATTTTCCA	1560
40	TCAATTAAAT GCTTCTTCGC TATTTGTCAG CCCACTTTTT TACCTGCAAC TTGTTAAATA	1620
	ATCCTTACAT CGTTAACGAA TAGTTCATCA TTTAGTTGAA TCAGCTCAAC TTTATTAACT	1680
	TCATATTTTC ACAAACATT GCGCAATCCA TTCCTTTTCC ACTACAAGCA CCATAATTAA	1740
45	ACAACAATTC AATAAAATAA GACTTGCAAA GCATAGTTAT GTAGCTATAT AAACGCCTGC	1800
	GACCAATAAA TCTTTTAAAC ATAACATAAT GCAAAAACAT CATTTAACAA TGCTAAAAAT	1860
50	GTCTCTTCAA TACATGTTGA TAGTAATTAA CTTTTAACGA ACAGTTAATT CGAAAACGCT	1920
	TACAAATGGA TTATTATATA TATGAACCTA AAATTAAATA GAAAGAAAGT GATTTCTATG	1980
55	ATTAAAAATA AAATATTAAC AGCAACTTTA GCAGTTGGTT TAATAGCCCC TTTAGCCAAT	2040

CmTyCAAArG AcACaGACAT TACTAGCCAA CGATTTAGCT ATnACTCCAA ACCTTCCATT 2160  
GGATTTGGTA AAGGnT 2176

5 (2) INFORMATION FOR SEQ ID NO: 446:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1557 base pairs  
(B) TYPE: nucleic acid  
10 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 446:

15 AAAAGCATGG CTTAAATGAA GTACGCTATA ACAAATTACA AGAACATGCT ATTGTTATGC 60  
ATCCGGCACC TGTGAATAGA GGAGTAGAAA TACAAAGCGA TTTAGTAGAA GCTTCAAAAT 120  
20 CAAGAATTTT TAAGCAAATG GAAAATGGCG TTTACTTAAG AATGGCAGTC ATTGATGAAT 180  
TATTAATAAATA GGTAAGGGGA CGAAAATGAT GAAATTAATT AAAAACGGTA AAGTATTACA 240  
AAATGGCGAA TTACAACAAG CAGATATTTT AATTGATGGT AAGGTAATTA AACAAATTGC 300  
25 ACCTGCAATT GAACCAAGCA ATGGTGTGTA CATCATAGAT GCGAAAGGTC ACTTTGTGTC 360  
ACCTGGATTT GTCGATGTTT ATGTTTCAATT ACGTGAACCT GGTGGTGAAT ATAAAGAGAC 420  
AATTGAAACT GGTACTAAAG CTGCTGCTAG AGGCGGATTT ACAACTGTAT GTCCAATGCC 480  
30 TAACACAAGA CCGGTACCAG ATTCTGTAGA ACATTTTGAA GCTTTACAAA AATTAATCGA 540  
TGACAATGCT CAAGTACGTG TATTACCTTA TGCTTCAATT ACAACACGTC AATTAGGTAA 600  
AGAATTGGTT GATTTCCAG CACTAGTAAA AGAAGGTGCC TTTGCGTTTA CAGATGACGG 660  
35 TGTAGGAGTA CAAACTGCAA GCATGATGTA TGAAGGCATG ATTGAAGCTG CAAAAGTAAA 720  
CAAAGCCATC GTAGCACACT GTGAAGATAA TTCATTAATC TATGGTGGTG CAATGCATGA 780  
AGGGAAACGC AGTAAAGAGT TAGGTATACC AGGTATTCCA AACATTTGTG AATCTGTTCA 840  
40 AATCGCAAGA GATGTACTAT TaGCTGAAGC AGCAGGTGTG CATTATCATG TATGTCATGT 900  
TTCTACTAAA GAAAGTGTTA GAGTCATTCG TGACGCTAAA CGCGCAGGCA TTCATGTTAC 960  
45 AGCTGAAGTT ACACCACACC ATTTATTGTT AACAGAAGAT GATATTCCTG GTAATAATGC 1020  
CATTTATAAA ATGAATCCAC CATTGAGAAG TACTGAAGAT AGAGAGGCTT TGTTAGAAGG 1080  
GTTACTAGAC GGTACAATTG ACTGTATCGC AACAGAcCAT GCACCACATG CACGTGATGA 1140  
50 AAAAGCACAA CCAATGGAAA AAGCaCCATT CGGAATTGTT GGTAAGTAAA CAGCATTCCC 1200  
ATTATTATAT ACGCATTTTG TAAAAATGG TGATTGGACA TTACAACAAT TAGTAGATTA 1260

55

TTATGCAGAT TTAACAATCA TTGATTTAGA TAGTGAACAA GAAATTAAAG GAGAAGATTT 1380  
 CTTATCAAAA GCAGATAATA CACCATTTAT CGGCTATAAA GTTTATGGAA ATCCGATCTT 1440  
 5 AACAAATGGTT GAAGGCGAAG TTAAATTTGa GGGGGATAAa TA-TATGCAA gCAAACGTTA 1500  
 TCTAGTGTTA GAAGACGGTC TTTTACGAGG CTACCGTTAG GTCTGATAAC TTA CTGA 1557

(2) INFORMATION FOR SEQ ID NO: 447:

10

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1799 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 447:

20 GCTAGAAATm TTGmATGaCA ATACAACTCT GTTAAaTGA TGGACGTAGA CAAATATGCG 60  
 TATTGACGCT TTATTTTAAA AATTaACATG CTTATAACAT GTTTATAGAA GGAGATTAAC 120  
 CTATGAACTA TCaAGTTCTT TTATATTATA AATATATGAC GATTGATGAC CtGAACAGTT 180  
 25 TGCTCAGGAT CACTTAGCCT TTTGTAAAGC ACACCATTTA AAAGGTAGAA TTCTTGTTTC 240  
 TACAGAAGGT ATTAACGGCA CATTATCTGG TACAAAAGAA GAAACCGAAC AATATATGGC 300  
 ACATATGCAT GCCGATGAAC GATTCAAAGA TATGGTGTTT AAAATTGATG AAGCTGAAGG 360  
 30 ACATGCTTTT AAGAAAATGC ATGTACGTCC TCGAAAAGAA ATCGTTGCTT TAGATTTAGA 420  
 AGATGACGTC GATCCAAGAC ACACAACTGG CCAATATTTA TCACCTGTAG AATTTAGAAA 480  
 AGCTCTTGAA GATGATGACA CAGTCATTAT TGATGCACGT AATGATTATG AATTTGATTT 540  
 35 AGGTCATTTT CGAGGTGCAA TTCGTCCAAA TATCACACGT TTTAGAGATT TGCCTGACTG 600  
 GATTAAAGAG AATAAAGCGT TATTTGCAGA TAAAAAAGTG GTTACGTA CTACTGGTGG 660  
 CATTTCGATGC GAAAAATTTT CTGGATGGCT TTTAAAAGAA GGTTTCGAAG ATGTAGCTCA 720  
 40 ACTTCATGGC GGTATTGCTA CATATGGTAA AGATCCTGAA ACAAAGGTG AATATTGGGA 780  
 CGGTAAAATG TACGTATTTG ATGACCGTAT CAGTGTTGAT ATCAACCAAG TTGAAAAnAC 840  
 AATTATTGGT AAGGATTGGT TTGATGGCAA ACCATGTGAA CGTTATATTA ATTGCGCTAA 900  
 45 CCCAGAATGT AATAAACAAA TATTAGTTTC TGAAGAAAAC GAACTAAAT ATTTAGGTGC 960  
 ATGCTCTTAT GAATGTGCTA AACATGAGCG TAATCGTTAT GTTCAAGCAA ATAATATTAG 1020  
 50 TGATAATGAG TGGCAACAAC GTTTAACAAA CTTTGATGAT TTACATCAAC ATGCTTAGTT 1080  
 TTAATTAAAT ACCTTTCAAA ACACGCTTTG AAAATCCGAT TTATAAAGGT TTTCAAGGC 1140

55

TAAATTTTAA TACTGCGGGG TGTCTTAAAA TGCACATTTT AGTAACAGGG TTTGCGCCTT 1260  
 TTGACAATCA AAATATCAAT CCCTCATGGG AAGCTGTGAC TCAACTAGAA GATATTATTG 1320  
 5 GCACACATAC AATCGATAAA TTAAAACTAC CAACCTCTTT TAAGAAAGTA GATAATATTA 1380  
 TAAATAAAAC GTTGGCATCT AATCATTATG ATGTTGTA CT AGCTATAGGA CAAGCTGGTG 1440  
 GTAGAAATGC CATTACCCCA GAACGTGTCG CCATTAAATAT TGATGATGCA CGTATTCCAG 1500  
 10 ATAATGATGA TTTTCAACCT ATTGATCAAG CCATTCACTT AGACGGTGCG CCAGCTTATT 1560  
 TTTCAAATTT ACCaGTTAAA GCAATGACTC AAAGTATTAT TAATCAAGGA CTTCTTGAG 1620  
 CACTTTCAAA TAGCGCAGGT ACATTTGTTT GTAATCACAC ACTTTATcAC TTAGGTTATT 1680  
 15 TACAAGATAA GCATTACCCT CACCTACGAT TCGGATTTAT TCaTGTGCCA TACATACCAG 1740  
 AGCAGGTcAT TGGTAAACCC GATACACCAT CTCATGnCCA TTGAGGAAAA GATnAGTTG 1799

20 (2) INFORMATION FOR SEQ ID NO: 448:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1341 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 448:  
 30 ACTTGGTTTT TTATTGTTTA TAAATAAAAC TCACTTAATA ATGTTTTCAT AATCTTCTTC 60  
 GACTACTTAA TTCTTTAAGA TATTCGTGAA AAGAGACATT ACACTAGTTA ATTTTCAAAC 120  
 AATACAAAAA GCGTCTACCT CCTACATATA ATTGTAGCGG AGATAGACGC TTAATATTTA 180  
 35 TTTAAAAATT ATTTTAAACC ACCGAATGTC ATAACATCAC GGGCAATCAT ACTTTCTTCA 240  
 TCTGTTGGAA TAACGACAAC TTTAAGTGGT GAATGAGGAT AGTTAATAAA TCCTTCTTTA 300  
 CCACGTAGTA AGTTTTTCATT TTTCTTAGGA TCCCAGTAAA CACCCATAAA TTCTAAGCCT 360  
 40 TCAAGAACTT TCGCACGAAT TTCTACTGAG TTTTCACCGA TACCTGCTGT AAATACGATA 420  
 ACATCAACAC CATGCATTCT CGCAGCATAT GATCCAATAT ATTTGTGAAT TTTAGAAGCA 480  
 AATACATCTA AAGCCATTTG TGAACGTGCT TTACCTGATT CAGCTTCTTC TGATAAGTCA 540  
 45 CGTAAATCAC TAGATGTACC TGATAATCCT AATAAACCTG ATTCTTTGTT TAAGATTTC 600  
 AATACTTGTT CAGCAGTTTT ACCTGTTTTT TCCATAATAA ATGGAATTAA AGCAGGGTCA 660  
 50 ATATTACCAG AACGAGTACC CATGTGTACA CCAGCAAGTG GTGTGAAGCC aTTGATGTAT 720  
 CAATAGATTT ACCGCCATCG ATAGCTGCAA TTGATGCTCC ATTACCAATG TGACATGAAA 780

55



TATGGCTTGT ACCATGGAAA CCATACTTAC GAATGCCATA ATCTTTATAA TAATGATATG 900  
 GCAAGCTATA TAGATATGCT TTTTCAGGCA TTGTTTGATG GAATGCTGTA TCAAAAATTG 960  
 5 CCACATGAGG GATATTTGGT AATAATTTAC GGAAAGCACG AATACCCATC AAGTTaGCTG 1020  
 GGTGTGaaG CGGTGCTAAT TcGCTTAATT CTTCAATTTT CTTTTCAACC TCATCAGTAA 1080  
 TAGCTACTGA TTCAGGGAAT TTTTCACCAC CATGTACAAC ACGGTGACCT GTTCCATCGA 1140  
 10 TATCGTTAAT ATCATTAAATA ATATTGTGCG CTTTAAAAGC ATCCAACATG ATATCAACTG 1200  
 CCTCAACGTG ATCCTTGATA TCTGTACTG TTTTAACTTT TTCCCCGTG ACTTCAATTG 1260  
 TAAAAATTGA ATCCTTCAAT CCGATTCTTT CTAATAAACC TTTTGTACT AATTCCTCTT 1320  
 15 CAGGCATTCT AATTAATTGA A 1341

(2) INFORMATION FOR SEQ ID NO: 449:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1529 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 449:

TTTTGAAGAT ACTACCGATG AAAATAGACA AAAGATTTTT CAATATTTAT CACCTGAaG 60  
 30 AGTTGCAAAT TTCTTTGATC AATTAGATAT TGATGACGAT GAATATGAGT TGCTATTTGA 120  
 TAAGATGAAT GCGACATACG CAAGTCACAT ATTAGAAGAA ATGTCATACG ACAATGCAGT 180  
 AGATATTTTA AATGAGTTGA CTAAACCAAA AGTTGCTAGT CTTTAAACAT TGATGAATAA 240  
 35 AGATGACGCG AATGAAATCA AAGCATTACT TCACTATGAT GAGGATACGG CCGGCGGTAT 300  
 TATGACGACG GAgtATTTAT CACTTAAAGC GCATACGCCT GTTAAAGAAG CATTATTATT 360  
 GGTCAAAGCG CAaGCACCAG ACGCAGAAAC AATATATGTT ATATTTGTCG TTGATGATGA 420  
 40 TGGTAAATTA GTAGGTGTTT TATCGCTAAG AGATTTAATT GTAGCTGAAA ATGATGCTTA 480  
 TATTGAAGAT ATTATGAATG AACGTGTCAT TAGTGTGAAT GTAGCAGACG ACCAAGAAGA 540  
 TGTTGCTCAA GTTATGAGAG ACTATGATTT CATGGCTGTA CCTGTTATAG ATTACCAAGA 600  
 45 ACATTTGCTT GGTATCATCA CGATTGATGA TATTTTAGAC GTTATGGATG AAGAGGCTAG 660  
 TGAAGACTAC TCTCGTTTAG CCGGGGTATC AGATATCGAT TCGACTAATG ATTCAATCAT 720  
 50 TAAAACAGCA TTA AACGTT TACCATGGTT GATTATTTTA ACATTTTATG GAATGATTAC 780  
 TCGACAATT TTAGGGAGAT TCGAAAAAAC ATTAGAAAAT GTAGCGCTAC TCGCAGCGTT 840

55

TCGTAACATT ACGACAGGGG AAATTAATGA GCAAAGTAAA TTTAGAATTG CATTAAGAGA 960  
 AGCAGGAAGT GGTGTATTAT CGGGTGTGTG ATGTTCAACA ATATTATTTA CAATTATTGT 1020  
 5 TGCAATATAT CATCAGCCAC TTTTAGCATT AATCGTTGCA GGAAGTTTAA CTTGTGCGAT 1080  
 GACGGTGGGG ACGTTTGTAG GTTCGATGAT TCCATTATTG ATGAATAAAT TAAATATCGA 1140  
 TCCAGCAGTG GCTAGTGGAC CATTTATTAC AACAAPTAAT GATATTATTA GTATGTTGAT 1200  
 10 TTATTTTGGT TTAGCTACAT CATTTATGGC TTACTTAATT TAAGGAGGAG TTATGGAGTT 1260  
 TTTATCTTTA GTTATTGTTG TTTAGCAGC GTTTTAACT CCAATAATTG TCAATCGATT 1320  
 AAATATTAAT TTCTTGCCAG TTGTTGTTGC AGAAATTTTG ATGGGGATTG TGATTGGAAA 1380  
 15 TTCATTCTA AATATAGTAG AAAGGGATTC AATTCTAAAT ATTTTATCAA CGTTAGGCTT 1440  
 TATCTTTTAA ATGTTTTTAA GTGGTTTAGA AATTGATTTT AAAGCTTTTA AAAAAGATAA 1500  
 ACGCGCACGT CAAGGACAAA ATGATGATG 1529  
 20

(2) INFORMATION FOR SEQ ID NO: 450:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1827 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 450:

TTCTGGAAAC CAAAGTATTG TCATCTTCTA CTAGTAGTAT nGGCATCCAT ATCACCCAAT 60  
 ATCATTTAGT ATATTTTATA TTTTCTCCTG ATTITAATCG ACTTTGAAAA TCTTTAATCC 120  
 35 GGCAGTCAAC TTCAAAGCCA TGAATCATCA TTTTTCATG TGGTGCTTGT ATAAAGTAAT 180  
 AAATCGGCCA AAATAATCGA GGGATATAAT CGTATAGATG TATATAAAG ACTGCCGACT 240  
 CTTTGATTAA TCTAAATTCT AACTTCCCTT GATTAAGTGT ATATTTTTTC ACTAACTTC 300  
 40 CACTCAATAA AATTAAAGTT ATTATTCCAT CAGCTGTTTC TTCTATTTTA AATACTGCGA 360  
 GCGGTCGCAC CTTATTCTTA ATATATATCT TAAATTGATC ATGTGATTTT TCTGTTTTCA 420  
 CAAAAGTTCC TTTAGTGTA CCCATCCATG CAATAAAATG GTTTACAACG TTCTTTAATG 480  
 45 TCCATCCCTT TGGTAAACT ACCTTCATCG TTGATCTAAC ATCATCATAC TTTGAACTT 540  
 GTAATTCTAC ATTAATAAAA GAACGTTTAA AACTAAATT TGTGTTTCT ACAGGTGTAC 600  
 50 CATATGCACC TAGGCGTTCA ATTGTTTCAT TATCATAACG ACTCCCAGGT ATGTAGATTA 660  
 CTTTTTTTAC TTGATTGATA GCCGCTGCTC GACCAAAATT ATCTGCTGCG ATTAATGTTA 720

55

CAGCAATATC TATTTGATTC ATTGCTGCAA CAACCTGTTC GTAATGAAAT ATATCACACT 840  
 GAATCCAAGT CATTTCAACA TCATCTGTTT KTTTATGTG TGGATATTTT GATATAGCAA 900  
 5 AAAGTTCAGC ATCATTTTCA ATCACTTCAC TTAAATACTT ACCAATATAT CCTGTTCCAC 960  
 CTGCTAATAA AACTTTAGGT TTCATCTAAA ATACTCCTTT AAAGTGTAA CAAAAACAT 1020  
 ATTACTCCAC CTTTGTAGTA CATATATATT ATAATAGTAG CAAATGTTTT AAAATTTCAA 1080  
 10 AATACTGGAG GCTTTTTATG GCCCATATTA TACGTAGAGT TAGTATCAA GATGTAGAAA 1140  
 ATTTCAATTC AATGTTAGCG AACATATACG ACGAATCTCC GTATATGTTC TACACACCAG 1200  
 GAGAATATGA TCCTAGCGTC ACATCGGCTA GTAAACAATT AGAAGAATAT ATCACTTCTC 1260  
 15 CGCATAAAGT CATCTTCGTT GCTGAAAGTG ATGAACAAC CTGTTGGCTTT GCCTTTGTTA 1320  
 ATACGACACC ATTTCAACGC ATTAAACATG TTGCTAAAAT TGATTTAGGT GTAAAGAAAT 1380  
 TATATCAACA TCGTGAATT GGCCAAGCAC TTCTTGATGC CATTATGGCT TGGTGTTTAA 1440  
 20 ACAATCAAAT ACACCGAATT GAAGCAAATG TACCACTCAA TAACCAACCT GCCCTCGAGC 1500  
 TTTTAAAAG TGCCGATTTT CAAATCGAAG GCGTTTTAAA AGATAAGTTA TTTATCGATG 1560  
 25 GTAAATATTA TGATGACTAT ATGATGGCTA AAATTCTTAA TTAAAGCTAT TTTATCATAA 1620  
 TCTGTATCA GAATCGTATA ACAACGAATT TAATGGTTAC CTAATACATT ACTCATACTT 1680  
 ATCAATGTTA TCTAATCTCA AATAAATACG TACACTCTTA TTCATTTATC AAATTTAAAT 1740  
 30 TCAAAATANA ACACCACTAA TGTGTAATTG ATTAATATC AACTACGATT AGTGGTGCTT 1800  
 TATATATGTG GTTAGTTTTT CTAATA 1827

## (2) INFORMATION FOR SEQ ID NO: 451:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 616 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 451:

45 ATATAGATTA ATGTTTGTTA TTTGTACTGT GTTTGCCAGT GACATTTATT ATAAACATA 60  
 TGACGTTATT CTATGKTGTA CTCATTTCTT ATATTGTAGG TTTATTAACG ATTAGAAGTA 120  
 TTATTAAAAA GTTGAAATAT CAGGAAACAT TATTACGAGA CTAAAAAACT TCCATTGGCA 180  
 50 TGTATGTTGT AAAGGTGCAT GTAATGTTGA ACGCCAAATG ATACGGCGTT CAGATTACAT 240  
 TAGCATCTAT ACGTTAACAG CATAACCAAT GGAAGTTTTT TTCGAATCTA TTCTTTTATT 300

55

AATGCGTCTT TTGAAAAATG GTCATTAAAG GCATCAGATT GCTTAAAGTC TTCGTATGCA 420  
 TGTCGATCAG CAAATCCGAA ATAAATTTTG TATGTTGTAC CTTTAGCAGG TCTTAACAAA 480  
 5 CGATAGCTTT TAAAGCCACC AAAGTTTCTG AAATTATCGT CTACACTAAT CAGTTTCTTT 540  
 TCAAGTTGAT ATGCATGATC TTCTGTTGAT GGaATGAaG A TTGCACaATA GAAATGaTGT 600  
 TCAcTGAATT CACCAA 616

10 (2) INFORMATION FOR SEQ ID NO: 452:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 944 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 452:

GCACGAGTGA TTAAACGGTT AATCAATGAA ACATTGTGATG CAAATTACAT TGAAGTTATT 60  
 GAGGGAGGAA TTGAAGAAAC GCAAACGTTA ATTCACCTAC CTTTGTACTA TGTCTTCTTT 120  
 25 ACAGGAAGTG nAAATGTAGG CAAAATCGTT TATCAAGCTG CCAGCGAAAA TTTAGTTCCT 180  
 GTGACATTAG AAATGGGCgG TAAAtCTCCa GTCaTCGtTG ATGAAACAGC GAATATTAAA 240  
 gTTGCTAGTG AGCGCATTTG TTTTGGGAAA TTCACTAATG CCgGCCaAAC ATGTGTTGCa 300  
 30 CCAGATTACA TTTTAGTACA CGAATCTGTA AAAGATGATT TAATCACAGC CCTATCAAAA 360  
 ACGTTGCGTG AATTTTATGG TCAAAATATA CAACAAAGTC CAGATTATGG CCGCATTGTA 420  
 AACCTTAAAC ATTATCATCG TCTGACTTCA TTA CTTAACA GTGCACAAAT GAATATTGTA 480  
 35 TTTGGTGGTC ATAGTGATGA GGATGAACGT TATATAGAAC CAACATTGTT AGATCACGTT 540  
 ACAAGTGATT CAGCAATTAT GCAAGAAGAA ATTTTGGTGC CTATCTTACC GATTTTAACG 600  
 TATCAGTCAT TGGATGAAGC AATAGCCTTT ATTCACCAA GACCAAACC TTTGAGTTTA 660  
 40 TATTTATTTA GCGAAGATGA AAATGCTACA CAACGTGTAA TAAACGAGCT ATCATTTGGC 720  
 GCGGCGCTA TTAATGATAC ATTGATGCAC CTAGCGAATC CTAAATTACC ATTTGGTGGT 780  
 GTTGGTGCCT CAGGTATGGG ACGCTATCAT GGTAATATT CATTGACAC TTTTACACAT 840  
 45 GAAAAAGCT ACATTTTCAA ATcAcACGA TTAGAATCAG GTGTCCATTT ACCACCATAT 900  
 AAAGGTAAAT TTAAATrCAT CAAAGCTTtC tTTAAAnATT AATT 944

50 (2) INFORMATION FOR SEQ ID NO: 453:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4820 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 453:

	TCTCCAGTAG ACCTTGTGTA TGAACAGTTT CTTTCATATG AATGAACATC GTTTTAAAG	60
10	TTTGTTCCTCA CTCAGCCTTA TCAGGATAAT ATCTAGAGAC AGTCGTCTCT GGCATCTCCA	120
	TTGTATGATA TTAAACCTTA TGCAGCGACC CATGATTTTC GTTATAAACA AATGTATGAT	180
	TCACCTTCATC GAAATCATGA TCTTCTCCTG CAATCCAAAA AACTGGTACT ACTTGTGCT	240
15	TATGTGTATC CGTTAATTCC TTAGATAAAG TAATGATTGa AAATATTTTA TGAATGTAT	300
	ACAATGGTCC CCCGAAAAGC CCTGCTTGTT GTCCACCAAT CACAACTTTT GAACCATTAG	360
	CTAAATGTTG TATGTTTAAT TCTTGTTTAC TTGAAAGCTT GAAAGCTTTA AATCACTCAT	420
20	ATATTCACGA ATAACATTCG CTAATGCCGC TTCTCTTCCA TTATTTTCTT TAGACATCCT	480
	TTTTTCAAAA CTAGTTTGTT GAGCTGCATC ATATTGAAAT AATCCTGTTA TTACAGGGTC	540
	ACTGCTCTTT ATTTTGGTA TAACTGATC TTTTTCATTT AACTAATA CTTTACAGTC	600
25	CATGTTTTTT CTCCTTAAGT ACGCGATTAC AACTATAGT ATAAAGTCTA TACCGGTGAT	660
	TGACAATTTT ACGGCTTGAA AATCAATTTA ATCATGGAAA ATTTATAATA TTCATTGTTT	720
30	TACATTTTCA AATCAATGAA AAACACAAGT GGTTAATGT ATAATAATAG TAGTAAACAA	780
	ATAAGGGGTA GATAAATATG AGTGAAATCA AACGCTTGA AATTAATTAC AAAACTGACG	840
	AATTATTCGA AAACCTTTAGA GCGTTTGGCA ACAAAGACTT ATACATGGTC AATGAGTTAA	900
35	ACGGTCAAAT GATTGATGCA AGTTCAGATT CACCATTTTA TGGCATATTT GTCGaGATCA	960
	ATTAGGAGCT AGAATGGCAT TACTAAAAAA AGGTGATGTC GAAGAAATCT ACTTCCCAGA	1020
	TTTTGAAGAT TATATATTAT TATGGAAGTT AGAAGTATTA CCAAAATATC AAAACAGAGG	1080
40	GTACGCTTCA GAATTGATTG ATTTTGCAA GAGTTTCAAT ATGCCAATTA AAGCCATTGG	1140
	CAGAAATGAT TCTAAGGATT TCTTTTACA TCATGGATTT ACAGATGTGG AAGCTAAAAA	1200
	TATAGAGGGA CATGATGTCT TATTGTGGAA ACCATAAGAT AATAATATTC GACACTACGA	1260
45	GCATGAAAAT GCATCTTTTC GTAGTGTCTT TTTTACAATT ACTTTCTTAA GCTAATATAA	1320
	GTaAATCATT TTCAAATTAT TTGTCTTAAC GTACAATATC ATTTAGTTGT TTCCATGrAT	1380
	TAATTTTATA ATCAGGTATA ATTCCTGGAT TATGATCAAA TCCTCTAAaA TTAAACCAGC	1440
50	AAGTAGCTAT ACCcGCATTG ATTCCACCTA GAATGTCAGA TGTTAnAGAA TCTCCaACTA	1500
	TAATCGAGTG CTGtCTTTCA TCCTCACCAA TATCATTAaA AACATAATTA AAAAATTCCG	1560

55

EP 0 786 519 A2

	ACGGCGTCTG ATTTAACCTT CTCTTTTGCG TTTCGGTTAC ACCATTAGTA ACAATATATA	1680
	AATCATGTCT TTTCGATAAT TCGACAATTG TTTCTAATGT TTGATCAAAG TATTTAACTT	1740
5	TAGCTTCTGC TAATCCATTT CTAAATAACA CATCTGCACG ATGCCCATCA ACTTCCATTT	1800
	GATGATGTTT GAAGTAATTC ACAAATCGTT CTGATAATAC TTCAGACTTC GkTAATTTAT	1860
	TTTGkTGAAA AGCTTCCCAA TGTGGTgAT TGaTTTTTTT AAATGkTAAA AAATCATCCy	1920
10	TTGTTGCTTT ATGATTAAAA ACATTCGCCA TATAGTGAAn CGCCCATCTCT TCTGcATCAT	1980
	AAAAATCAAC AATTGTATCA TCAAAGTCTA TCAAAATATT TTTATATCCC AATTTCCCCA	2040
15	TCTCCTATAT TGTCTATGTA TCTAAATCTT AACAGAGGCT CAAATTTCTG CAAATAAAAT	2100
	AAACTGAGTG CATAACATTA AAGTATGCTC ACCCAGTTTA TTTTAAAGAA TATTAGTTAT	2160
	TATATTAGAA TCCAAATAAT TTACCTAGTA AACCACACC GTTAGCAACG ATGTCTACGA	2220
20	TACTTGCGCC TAATTTcACA CTATCATGTT GTTGTGCAGC TTGCACAGTA TTTGCGATTG	2280
	CTTCTGCTAG TCCAGTCATT TAAATCTCTC CCTCACCTTT GAAATAATAC TGATTACTTA	2340
	CATAACATAT TGAAATTAGA ATCCGAATAA TTTACCTAAT AAACCTACGC CATTTTCAAC	2400
25	GATGCTCACA ATGCTTGTGC CTAATTTTGC GCCATCATTA TTAATTGCTG CAGTTACGGT	2460
	ATCTTTAATT GCGTTAAATA AACCTTCCAT TGAAAACACT CCTTAAATTT TAAATTTGAA	2520
	GATAACAAAA ACGTGCGTAg yTTTTAAATC ACCGAAATGT TATTCGCTTA ACGTTTTGTT	2580
30	GTGTTATTT TAAAATAAAT TTGATGCAAT TAGTTTGT TT ATCCGCACAA CATCTTATAA	2640
	TGTACTTAAC TGTATTTTAA AGAGAAAAGA AATACAGTTA GGCATTCAAA ACTGTATTTA	2700
	ACACAATTAA GTTGCCCTGAA TTCGTATTTA AGTCTTATG AACCTTTTTA GATAAATAGC	2760
35	TCTATAATAG TGAAAAATAT AAACATTTTT TATTTACAAG GTATTGCTAA TTAAAGTTCA	2820
	TTTAGATATA ATAATCTTG TGTGTAAAA CGTGTCTGCG TAGCTCAGCT GGATAGAGCA	2880
	ATGGCCTTCT AAGCCATCGG TCGGGGGTTC GAATCCCTCC CAGGACGTTT ATAGGTATTT	2940
40	TTATACGCAT TACCAAACAA AAGAGTTCCG TGATTACGGG GCTCTTTTTG TTTTGAATTT	3000
	CAGTAATATA GTATGATGCG TCACCAAAC GTCCCCGCA TAAGCCCCGA AAATACAGTA	3060
45	ATTAAAACAA GCATGCTTAT TCGTTATAGA ATTTTTTGAC ACACAATTGA CACGCGTCTG	3120
	ACACTTGTTT ATACATTTTT AATTAAGTAA TTTTGTGCTC AAATTCATC TATACTGCAC	3180
	CTGAACTACA CCAACACTAC ACCAAGATTT TTAACACTCA CCATTGTCAT GCGTAGAGAT	3240
50	TTTATTATT ATATTATTCC TATAGATTTT GATACTATTC AAAATTTTAG GGACTTTTCA	3300
	GGGGCCCGAA ATCCTATAAT TATAATTATA TACATCTAAA AAAAATAACC ACGTCCATCG	3360

55

5 ATACTATTGG CAAATTTATA AAGTAGTTCA GCGTTTTTCA ATGACATATT GTCTAATGAT 3480  
 CTTTCATTTT TTCTCATTCT GTGTATTGTG CTTTGTGGAA CTCCTGTTTG TTTCGATATA 3540  
 10 TGTAAACTGC TCAAATCACT GTCTAATAGT TTTTGAATTT GATTTCTCAT TTTGTACACC 3600  
 CCCTTGTAAG TCTTCAATCA TCATTTAAAT TAAATAATTA CTTTTCACA TATTCCAATA 3660  
 TTTAGGTTGC AAAGCATACC TCAAATATCA TTAATTTTGA GATTTAAATG TCAATTATGT 3720  
 15 GTCTATCAAT CCAATATACA TACTCTAATA ACGTAATAGT ACACACTCTT CTTATTAAAT 3780  
 GGCCATAGCT ATCATGATAT AATTAGTGAA GAAAATCACA TAAGAAAGGT TGTAATCAT 3840  
 GAGACTTCAA AAAGCACCTC TAGTAACGTC AGGACTAGTC TTAGGATTAT TAGGCCTGGG 3900  
 20 TAATCTATTA AAAGACTTAT CTCTTACTTT AAACGCTGTT TGCGGAATCT TTGCTTTCTT 3960  
 GATTTGGATT CACCTTTTAT GTACTATGAT CAAATATTTT AATAATGTGA AAGAACAATT 4020  
 AAACAGTCCT CTAGTTTCAT CAGTGTTCAC AACATTTTTC ATGTCTGGCT TTTTAGGTAC 4080  
 TACTTATTTA AATACATTTT TTAGTAACAT AACTTTTATC AATAGCTTAA TAACGCCTAT 4140  
 TTGGATTTTA TGCCTTGTGG GAATTATGAC GCATATGATT ATTTTTTCAA TAAAATATTT 4200  
 25 AAAAGATTTT TCACCTGAAA ATGTTTATCC TTCGTGGACT GTACTTTTTA TTGGTATTGn 4260  
 TATCGCAGGA TTGACGGCAC CCGTTAGCGG ATATTTTTTC ATAGGTCAAT TAACAGTAAT 4320  
 ATATGGCTTT GTAGCTACTT GTATTGTCTT ACCTATAGTT TTCAAGCGAT TAAAAGCATT 4380  
 30 TCCATTGCAG ACGTCAATCA AACCGAACAC ATCGACAATT TGTGCACCAT TTTCTTTAGy 4440  
 CGCTGCAGCA TATGTTATAG CTTTTCCTAA GCGGAATGCT TTTATCGTAA TTATATTTTT 4500  
 ACTATTAGCT CAAATATTTT ATTTTATAT CATTATACAA TTGCCTAAAT TACTAAAAGA 4560  
 35 ACCTTTTTTCG CCCGTATTTT CAGCTTTCAC ATTCCCTTTA GTAATCTCAG CAACTGCTTT 4620  
 AAAGAACAGT TTGCCTGTAC TTATGTkTCC AGACATTtGG AAAGGkCTTT TGTTTATCGA 4680  
 AGTGTTATTA GCCACTGTAA TAGTACTTAG AGTCTTTATA GGATATCTTC ACTTCTTTTT 4740  
 40 AAAAAAGGAA AAACAAGATA AATTCTnCG TAATGCGTCT CAGTAACACT ATTACCAAGA 4800  
 ATTAACACGT ATATTTAATA 4820

(2) INFORMATION FOR SEQ ID NO: 454:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4358 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

	ATTTGGATCT TTAATATCAC CAATATTTTT AATATCTTCC GGATTCAATC CATATACTTG	60
	TACTGTATCT GAGTATTTAA TTGTGAAATA ATCACCTGAT TTAACCTTGT CATCAACTGT	120
5	AATTTGTGAT TTTAATGATA AATAATCTTG GGCTGGTACG ATTTTATGT TTTTATCTGC	180
	ATCAACGACA GTtAATGTTG TATTGATGT GATTAAATCA TTAACATTTT TAGCCTCTGT	240
	TGATGATGGC TGTACTGCTG CTATACGCAT TCTGTATTG AAACGTTTAG GTGCTGTACT	300
10	TTTTGGCAAA ATGATATCTG CATTATTTTC ATTATTTGAA TTACTATTGT TATCAACAAG	360
	AGTTTCATCA TTACTCTTGA TAGCATCACT TTTAACATTT AATGTAGTTG ATTCAGTTT	420
15	GGCATCTACC TTTTGTGTTT CCTCATTAGT TGGTTGAACA TTTACCACTG ATTTATTCTC	480
	TTGCAATCA GGTtGTAAAG CTTCTTGATT ACTTATAGTT TGTtTAGTGT TTAAATCTTC	540
	ATTCGTAGAT TTTGGTGAAG CTTGCTCATC TGATTTGGCA GTTGAACTT CAACTTTATT	600
20	TCCAGTGGTA GATTGTACAC TTTCTTTTTC TATTAATTTA TTCCCATTTG AAGTCGTTTC	660
	ATTACCTTga GATGATACCA TTTCTTTTTC ATTATCATTT TTAGTATTGT CTTCTTGATT	720
	TAGTTGCTGC ATATCAACTT TATCACTCGA TTGATTATCA CTTGCTGAAG TTGTCGCTcG	780
25	TTCAATTCTT TATTAGTACT TTCTGCAGCC TTTGCTTCTT GGTTCCCCAG ACCAAAAATT	840
	AATGTTGTAC CTAATAAAT TGATGCTGTT CCCACTGTGT ACTTTCTAAT CGAAAATTTA	900
	TTTAATCGAT TGGATACCAT GCCTTTCCTT GTTATTGCCG TTTTATTTTC TCTGTTTAGC	960
30	ATTAGATTAC TCCTAATTCA TCAAATTTTT AAATAATACA ATTGTTTTAA ATACAAAAAT	1020
	GTATATCAAT ATAGTATTAC ATTTTTAGAT AAAGCACAAT ACTTTAATTA TTTTCTTTA	1080
35	TCGTAAAACG TTATTTAACA TTTGTGTTTA AATAAAAGTT TTTATGAGTt TTGTAATCTT	1140
	TATTTAATCA TCATAAAAAA TAGTATTATT TGCCCTTGAA ATTAATATCT TAGCTTTTCT	1200
	AATTCATAGA CAATTACATT TCTGTAACAA ATTAAATTGT ATCTATTCCT TAAAGATTTT	1260
40	TTGTTTTATA TCTGGGAATT TCTAAACAGA AAAAACCAGG CCACATGGAC CTGGTTAAGT	1320
	TAATCATATT ATTTATTTTG TTTTTTACGA CGACCGAATA ACAATAATGA TCCTAATGCC	1380
	GCGAATAATC CACCGAATAA TGTGCCATTA TTTGAATTAT TATTTTCACT ACCTGTTTCT	1440
45	GGTAATGCTT TAGCTGTTTT ATGCTGATCT TTAACCGTAC TCATTGGTTT AGCCGGAGTA	1500
	TGTTTACCTG CATCTGAATC TGAATCGCTA TCTGAATCTG AGTCGTTGTC TGAGTCCGAA	1560
	TCGCTATCTG AATCTGAGTC GCTGTCTGAA TCTGAATCGC TATCCGAGTC TGAGTCGCTA	1620
50	TCTGAGTCTG AGTCGCTATC TGAATCTGAA TCGCTGTCTG AGTCTGAATC GCTATCTGAG	1680
	TCTGAATCGC TGTCCGAATC TGAGTCGCTA TCTGAATCTG AATCGCTATC TGAATCTGAG	1740
55		



	TCTGAATCTG AGTCGCTGTC TGAATCTGAA TCACTGTCTG AGTCTGAGTC GCTGTCTGAG	1860
	TCTGAATCGC TGTCAGAATC TGAGTCGCTA TCTGAGTCTG AATCTGAATC ACTGTCTGAG	1920
5	TCCGAATCGC TATCTGAATC TGAATCGCTA TCTGAGTCTG AGTCGCTATC CGAATCTGAG	1980
	TCGCTATCTG AGTCTGAGTC GCTATCCGAG TCTGAATCGC TGTCTGAGTC TGAGTCGCTG	2040
10	TCTGAATCTG AATCGCTATC TGAGTCTGAG TCGCTGTCTG AATCGCTGTC TGAATCTGAG	2100
	TCGCTATCTG AATCTGAGTC GCTATCTGAG TCTGAATCGC TGTCAGAATC TGAGTCGCTA	2160
	TCTGATGTTT CTTCTTCGTA GTAGCCATTA TCAAGTGTGA AATCATCATG ATCCGTAATT	2220
15	GTTACATCAA CTTCCGCCACC ATCGGCATCT TTATCATCTT CAGTTGTATT TGTACCTGTT	2280
	TGAGTTAAGC CAGCAGGTTT TTCAAAGATA ACTTTGTATT TACCACTATC TAAATTATCA	2340
	AAGCGGTATT TACCATTTTC ATCTGTYTCA GTTGTAACAA TTACTTCGCC TTTTTCGTTT	2400
20	TGCAAAGTAA CTTTAACACC TTTAATTCCT TTTTCAGTCG AATCTTGTTT ACCATCTTTA	2460
	TTACTGTCGT ACCAAACATA ATCACCTAAA CTATATTTTG GTGTTTGTGA GAATCCACTA	2520
	TCTAATGTCA TGTGTGCAGC GTCTTTAATG ACACCTGTTG TAGTTAGTCC ATCAGAATCT	2580
25	ACAGCATCAT CTGTACCTAC ATTTGCAGTT GTCGGTGTAT AACCGGCTGG TGTGAAAAC	2640
	TCTACACTAT AAGTTCCATT GCTTAAACCA GTGAACTGAT ATTTACCATT TTCATCTGTT	2700
	GTCGTACGAT CTAATTCTTT ACCGTTACTA TCTTTAAGAA TGACATAAAC ACCTTTAATC	2760
30	CCTTTTTCAT TGGCATCTTG TTTACCATCT TTATTTGTAT CTTCCCATAC ATAGTCACCT	2820
	AGATTATATT TCTTTTGGTC GCCATTAGCA GTTGATGAGC CATTACATT TGAATAACTA	2880
35	TTTGACCAAC TATATTTAGT TTTGTCAGTG TCTAAAGTAT AATCAATTTT TCCATTATCT	2940
	GTTGAACTAT TATCTGGATA AGCAACTTGT TGAATGATGT ATTGTTTATT GCTGCTTGTT	3000
	TGGCCTTTCA TTAAATCGAC TGTAGCTGTT TTATTATCAT TACTATAAAT AACATCGAAT	3060
40	TGATCAGTAA CATCTTTAAG TTTTGAAGTA TCAGGGGTGA AACTATCCAC AAATTGATTT	3120
	TGATCTGtCA CtCGTAAAT TTTGAAGTTT TTTGCATTTG GATTAAATTT ATATCCAGTT	3180
	AAATTAGTAA CAAACGTTTG TTTAGTATAT GTATTTTtag GTTGATTTAC ATATGCAGTC	3240
45	ATATTACGCG ATAAATCTTC ATTGTTAATA TAGTTTGTAC TTGAAATAAG CGGTTGTGCT	3300
	TTTTTATTAC CATAATCGAC AATGATTTCT TCGCTATATG TATCATTACC TAAAGTTACT	3360
	TCCATTTTAT AAGCTGTTTT ATCAGTTGTT GCATTTTAC GTTTCGCAA TGCAACTTGT	3420
50	TCAAAGCTAC CTCTAACATT TGTATATTGA TCTACATAGT TCGTAAAAGT ATATGTTGTT	3480
	GTGTTTGTG TACTATCATA AATACCTTTT GCAATAATAT TACCTGGGC ATTATATAAA	3540
55		

GTAAATGTAT CGCCCTCTTT AACAGAATCA TCGATTGTGT AATTTGCTTT TAATTTTAAA 3660  
 ACATCACTTG AAGTTGCCCA AAATTCAGTT TTACCAGTAG TCTGATTAAC ATGTCCTTTA 3720  
 5 TCAATCGCAA TGTCAATATT TGAAAAATGT ACTTTATCAT TAACATTTGT TCCTTGTTGT 3780  
 GGAGCTGCAA CAGTATTCAC TGCCATGCGA TTAAAGTTC TTGGTTTAAT AGTCGTTGTT 3840  
 TTAGGTGTAG TTGAAACATC TTTTGCTTGT GTTAAATTAC TTTTATCAGT TTCATTACTA 3900  
 10 TATGTAGTTG ATGATTTATC ATTTGTTGTT ACATTGCTAG TTTTGTAGT AGATTGATTA 3960  
 GCTGTAGCGT TTTGTGGTGA TTGCATGTTA CTACTAGTTT CTTTAACTGT TGCACATCA 4020  
 CTCATTGTCA CTTTAGGCTG ATCTGCAGTT GCAGTTTTCG TATTGTCCTT TAGTTGACGA 4080  
 15 CTATCAACTT TTTTAGTTGT TTTATTCTCA CTGGGGCTG TCGTTTCATT TTTTGATTGA 4140  
 TTTAATCTC CATTCGTATG TTCTGCCGCT TTAGCTTCAT GACCACTTAA CCCAAAAATC 4200  
 20 AATGTTGTCC CTACTAAAAT TGAAGCAGTA CCTACAGAAT ACTTTCTTAT CGAAAATTTG 4260  
 TTTAATCGAT TTGGTATCAT GCCTTTnCTA TTGTnGCTG TCTTTTTATA ATTCATTTAA 4320  
 TAATACTCCT TTAAAATATC AAAATTTGAT AAATATAA 4358

25 (2) INFORMATION FOR SEQ ID NO: 455:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1060 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 455:

TTGACTTCTT AATTCAGCAT TTTCTGCACT TAATGCTTTG TTCTTTTTTAA TAAGTTGCTT 60  
 TCTTGCATAA ACTTCGGTAT CTATTTTACT ATTACTATAC CTTTGATTTA AACTAATAT 120  
 40 ACCAATTAAT GCTACAATGA TAATGATAAG TACAACATAA AAAGACATTT TTTCACCAAT 180  
 CCTTTTGGAC TTCTTTAACT TTGTATACAA TAATAATTAA TAAAGATTAA TTGTTATTCA 240  
 ATTTCCACA TTTTATTAG TTGATTTTAG TTCATCATTG TTATAATCAA ATTATAAACT 300  
 45 GACAGATATT GATGTTCAAT GAATATGACG TGAAAGATTG GTGAATTCAA GTTTATGTGG 360  
 AATTTATGTT ATAACGGTCA TTAAATGAC AGAATTAGGT CACTCATAGT ATTTTGAAGA 420  
 TTGAATTCAT TAATTTTAAA ATGTATAATG ATATTTGTGA AAGCGCTTGC TTAGGAGGTG 480  
 50 TATTTGAGAG TGAATGAAAT GAATGCTAAA GAACAATTAG TGGACAATTT AATGAAAACA 540  
 TCATCGCAAT TATTTAAATT TCACGGTGAA GTTGCCATGC AGCTTTTCTT AAATGATGAA 600

55

AAAGTTATTC CGCAATCATA TGC GTTACTA TACATAGATA AGCAAGATCA AGCAATAGCT 720  
 AAAGAAGATT TATCACTTTC AAAAATTGCA AAAGTTTATG TGCAATATGA TGATACAACA 780  
 5 ATAATGAGTA TTTTCGTTTA TGATGTAGTA AACGATGAAT GGATTTTATG ATTGGATCCG 840  
 AATATACGTA TACCTAAGAG TAACATATAC TTCCATAGTT TAAATTGGGA TGTGGATATA 900  
 TTAAACCGGA GtCGTCTAAT GTATGTCTAA TGCaCACCAT CAGaTCATCA TtATCCATTA 960  
 10 TAACGrGCAT AGTGCATAAn yACTwCatTT TaTTaaATTG AGaGGgGCAC GATAGGTGCA 1020  
 TCAGGACATA ATATAGGAAG CATCAACGCG TGAnCAGGTC 1060

## (2) INFORMATION FOR SEQ ID NO: 456:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1262 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 456:

25 ATAATnACTA AATACnAAAG TTAACTGTC TTACTAATAA TGA CTATGTT ATAATTTTAA 60  
 AAGTGATATT TTGGGTAATC GCTATATTAT ATAGAGGAAA GTCCATGCTC ACACAGTCTG 120  
 AGATGATTGT AGTGTTCGTG CTTGATGAAA CAATAAATCA AGGCATTAAT TTGACGGCAA 180  
 30 TGAAATATCC TAAGTCTTTC GATATGGATA GAGTAATTTG AAAGTGCCAC AGTGACGTAG 240  
 CTTTTATAGA AATATAAAAG GTGGAACGCG GTAAACCCCT CGAGTGAGCA ATCCAAATTT 300  
 GGTAGGAGCA CTTGTTTAAC GGAATTCAAC GTATAAACGA GACACACTTC GCGAAATGAA 360  
 35 GTGGTGTAGA CAGATGGTTA TCACCTGAGT ACCAGTGTGA CTAGTGCACG TGATGAGTAC 420  
 GATGGAACAG AACATGGCTT ATAGAAATAT CACTACTAGT TTAGCTCTCC TAGATGATGG 480  
 AGAGCTTTTT TCATGAAAAG AACACTTAAA ATTAACGCCy TGTCTTGaTA tAATGACaCT 540  
 40 GCcTTGTTTT AAAATAGTAA GCGGATGCgT TAATGTATCA GCGATTAAAT TTGTTGGAAA 600  
 TGTATAAAAA ACACAAGCTA AGAATAAAAT ACCTGTATAA AAGGAGAATC ATATATGTTT 660  
 CAATTACTTG CAGTTTGTC GATGGGATTA GAAGCTGTTG TTGCTAGGGA AATTCAAGAA 720  
 45 TTAGGCTATG AAACAAATGT TGAAAATGGT CGTATATTTT TTGAaGGAGA CGCAAGTGCA 780  
 ATTGTAAAGG CAAATTTATG GTTGCGCACA GCAGACCGAA TCAAAaTTGT TGTGGACGT 840  
 50 TTTAACGCAA CAACGTTTGA CGAATTATTC GAACAAACCA AAGCGCTCCC TTGGGAATCT 900  
 ATAATTGATA AAGAGGGTAA CTTCCAGTT CAAGGTAGAA GCGTTAAATC AACACTACAT 960

55

TATAACGAAA AAGGTTGGTT AAATGAATCA GGTGCCAAAT ACCCTGTTGA AGTTGCCATT 1080  
 TTAAAAGATA ATGTATTATT GACTATCGAC ACATCAGGTT CTGGTTTGAA CAGACGTGGT 1140  
 5 TATAGATTAG CACAAGGTGA AGCACCAATT AAAGAAACGT TGGCAGCAAG TTTAATCCGT 1200  
 CTTGCCAAAC TGGGAAAGGT GATTACACCT TTAATTGGT CCCATTTGCG GTTCnGGTTA 1260  
 CA 1262

10 (2) INFORMATION FOR SEQ ID NO: 457:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1142 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 457:

CCTGGCTGCT TTATCAGCAT CTACTACTTT AAAACCGAAT ACGGATAAGA GTTCTGATAC 60  
 TGTTGATTTT CTGAGGCGA TTCCACCTGT TAGACCAATA ACTTTCGGCA TAATTTCACT 120  
 25 CTTTCTTTAT TTTTGACATA CTGGACAATA ATGACTATTT CTTGTCGCGA TGATTTTGT 180  
 TTCAATTTGA CTTCCACACA CTTTGCATAC CGGCTGCTTA TATACATTAA GATGCAATTG 240  
 CATCTCACCA GTTTTTCCAT CAGCATGACG ATAATCTGAA ATACTTGTAC CGCCATATTT 300  
 30 AATACCTTCT TCTAGTACTT CTCTAACATA ATAAAAAACC ATTTCTTGTT GTTGGTGTGT 360  
 TAAGTCTTTT ACTTTTTTAT CTGGTAAAAC ACCTGCACGA AACAACGCTT CACATGCGTA 420  
 AATATTTCCA CAACCTGCGA TTACTTTATG ATCCAAATC ACTTGTTTGA TTGGTTTATT 480  
 35 CTTATTAGAC TGTTGATGAA TTCGATTTAA ATAATACGTC AATGCTTCAT TTGAAAAGG 540  
 TTCAGGCGCT ATTTCTAAAA ATGAAGGATA AGATGCTACA GACGCAACAT TTCTAATTTT 600  
 TCCAAAACGA CGTATATCTG AATAAATTAA CTTTTTGTCA TTTGACAACCT CAAAAATAAC 660  
 40 ATGCCAATGC TTACGATAAT TAGGTATCAT AATATCTTCA AGTTCATCTA CAATGAAAAA 720  
 ACCGCCCGCC ATACCTAAAT GACTAATTAA TGTACGTTGT TCTCGTTTAT TATCTAGCTG 780  
 45 AAAAACGATA TATTTACTTC TTCGTTCTAC ATTTGTAATG GTATAGCCTT CCGATAAAGT 840  
 TTTAAAAGTA TCTAATTCAA TTCCTTTTAT AATTGTTTCC TTGCCTTGAG CTTTACCCTT 900  
 GATTACTTTA TCCGAAAATA TAACGTGTTT AATTTTTTGA TTTATAACGT AGGGTTCAAT 960  
 50 TCCTCTTTT ACATGTTCTA CTTCTGGTAA TTCGGGCATA CCATTAACT CACTTTATTT 1020  
 TGCATCATAC CAGGTTGCAC CATAACTTGA GTCTACTTTT AATGGAACAT CTArTTGCAA 1080

55

TT

1142

(2) INFORMATION FOR SEQ ID NO: 458:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1814 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 458:

5	CCTTTAGTAA ACAATCCTTC TTTAGTTTTA GTACGTTGTT CCAATCCAAA TAATTTGTAT	60
15	TTCAATGCCT CGCCCGATTG AGTGCCGCTA AAGTTATCAT CTTTCATGTT AGGCGTGTGG	120
	GTAACATGT GTATATCACT GTTTAAACGG TCTTTATAAG CTTCGGTACC TTGTACATCG	180
20	TATTGCTTAT AAATATAACC ACCATCAACA GAGCCTTCTG TTTCTCTACC TTCGCTATCA	240
	GCATAACAG TCGGTTCTAA AAACAACACG TTAGCTTCCT TTTGTTTTCT AACTTCTACA	300
	GGATCTAAAT TTAAATTACC TTTAATAAGT AACATAGCGT CATTTAAATC ACTCATATAG	360
25	TTAGCAsymy CTGATTCAGC ATTATCATAC AAATCAATTA AAGTGATTAC TTTCTCATAA	420
	TCCCTTTTC TTCTTTCGTT GTTGCTAAAT TCTGTAATAG GCATACGTTT GAAAGAGTGT	480
	GATTCAAAAC CGTTTTTCAG TGGTGTGAGC TTCAATCCAT TTGTTCTACT GGTAAGATAT	540
30	CTATAAACAC CGTGTGAAGT GAATAAATCA ACTGTAAACA CTTCATCTTC GTCAGTCTTG	600
	TCTATTGGTT TAGTTCTTAA ATATCTAACG CCTGCGATAC TATTACGTTT AATTGTATTG	660
	TCGTATATGA CAAAAGTACT CATTGCATCA CTCTGTATA AACGCGTTTC ATCATCTTGG	720
35	TTTCTAATCA TTAATCATA AGCTTTGCCA TAAATTGACA AATCTAATCC TAAAGATCTA	780
	TTGTGTGACT CAACATCATT TAAATCATG AACGCCTCAA TAACTTCTAA TACATCTTGG	840
40	TCATCATCTT GATATTGAAT TGGATTACCC AAGAAATAGC CGTTGATAAA ATCGCTAATA	900
	TAAGATGCGT AATCATGCGC TACACGGTTA TCTGCCATGT ACTCTTCTTT GCGTCGTGTT	960
	AACTCAACTA AGTTCTTAGT TTTACCTTCG TAATAATCAC TTAACACTTT CAATCTAGGT	1020
45	CGTTGGTAAT CCATGTGATG TTCAATGTAT TTACTTACTT CATTAACGTT TTGTAATAAA	1080
	TCGGATTCCG TCCCGTCATA TGTGTAAACA ACATTGGCTT CATCATTAAA TAAGTAATTT	1140
	ATGTTTCCCC GTAGATCTGT ATCTGTTTCA AATTCGTTTA CTTTAAACAT TTGTTCCCTC	1200
50	CTATAATCCT AGAGATTTTA TTGTGTCAAC TTTCGAACTG AGATTGTGTC GTTTtCTAAC	1260
	CGGTCTGTAG AATCGTTCCA CTGAATAACG CAACGAATCG ATACAATGAT TGTATGTATC	1320

55

CTCTTCAATA GTCTTGAAAC AACGTTTCATC AACAAATGATT TCAAATTGCA TTAAGAATTG 1440  
 TAACCCTTGT ACAACCGAGC CCTTCCCTTT TTTGGTTGGT AAAATCCTTT TAAGCCCTAG 1500  
 5 ATTCTTAAT TCAGCTATAC TTTTGTGTTG TGCATATCT GCTGTAATTT CTCTTTTAGC 1560  
 ATAACCAAGT TGCTTTATGA CATTAGCTAT TTCATCATTC AGCATACCTT GTTTAACATA 1620  
 CTCTTCAATG ATGTATAACT TCTTTTCTT TACATCTATT TTAGAATGTA TAAAAGCACT 1680  
 10 AGGATCATTA ACGTAGCCAA AGTCCAATCC AAAATAAGAA GGTAAATGTC TTAATCATC 1740  
 TTTATTTATT AAACGTTTTT CATACTTAGG GAAAACCAAT TTGTCTAGTG TAGCAAATTC 1800  
 ACCTAACGCA TAAA 1814

## (2) INFORMATION FOR SEQ ID NO: 459:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 686 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 459:

AATTnAGATT ATTACCCTCC TTTAAAAATA TTTGTTTCAC AATTTTTTAT TACCTATTTA 60  
 CTGGTTTCAT GTCTTATGGG GCATTTTACT TGTGTTGATT TGAAAATGTG CAAAATTTAA 120  
 30 TCTTATATGT TTCTTGGCTT TTCATGACTA TGCTATTTAT GTTTATGAAT ATGCATTCAA 180  
 TTATAGATAA AAAAGTACAT ATATTCTTAA AGTCTAATAA ATAGTTACAA ATTTAGTTAG 240  
 TTTTCAATTG TTAATTAGGG GTGGTAAACA GTGCTTTGTG AATCTAGACA AATTATATAA 300  
 35 AATCCTAAAT ATCGAGTTAT TAGATATAAT AATGAATATT TCATGGTCGA TTTAGTAAGT 360  
 ACTTGGATTA CTTATTTTTT CCCTATGATT AATTGGTTTT TGCCCAAAA ATACGCAAAA 420  
 ATTAGCGAAA ATGAATTGA AAGGTAAAT ATAGTCGAGC CTGTTAAAA TAATGTTTTT 480  
 40 TGGCCGGTTG CAGGAAGTTC AGTTCTATTT GGAATTATAT TGAGAAAGTA CGGTAACCTC 540  
 TTTAATGTTT AGTTTGAAAA ACAACTAGCA ATCACTGTAT TTTTATCAT GTTAATAGGG 600  
 45 ATGTTAATTT TTTATTTTTA TCTAAATAAA AAATTAACAT TAAAAATTT TAATACCAAC 660  
 GTGGGTAATA AGAATAGGAG TTGTAT 686

## (2) INFORMATION FOR SEQ ID NO: 460:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1300 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 460:

5	ATCTGCAATT ATGGGCACAC CCAAGCTnAT GCAAGTAACT AAAGGAGAAG TACTTTTAGA	60
	CGGTGTAAAT ATTTTAGAaT TAGAAGTTGA TGAAAGAGCA AAAGCAGGAT TATTCTTGGC	120
	AATGCAATAT CCATCAGAAA TTACAGGTGT TACAAATGCT GATTTCATGC GTTCAGCAAT	180
10	CAATGCGAAA CGTGAAGAAG GACAAGAAAT CAACTTAATG CAATTTATTA AGAAATTAGA	240
	TAAAAACATG GATTTTCTAG ACATAGATAA AGACATGGCA CAACGTTATT TAAATGAAGG	300
15	TTTCTCAGGT GGAGAGAAGA AACGTAACGA AATCTTACAA TTAATGATGT TAGAACCTAA	360
	GTTTGCaATC TTAGATGAAA TCGATTCAGG GTTAGACATC GATGCATTAA AAGTTGTATC	420
	TAAAGGTATT AACCAAATGC GTGGGAAAAA CTTTGGTGCA TTAATGATTA CACACTATCA	480
20	ACGATTATTA AATTACATTA CTCCTGATAA AGTACATGTA ATGTATGCTG GTAAAGTCGT	540
	TAAATCTGGT GGTCCAGAAT TAGCAAAACG TCTTGAAGAA GAAGGATATG AATGGGTTAA	600
	AGAAGAGTTC GGTTCAGCTG AATAATCTTA TTAATACAGT ATCCATGAGA TGTTTCATCTA	660
25	TATATGATGA AAATGAACAT TTATACGAAA TAGTAAATTT CATCAAGTAG GAGGAAAAAG	720
	TTATGACAAC TGATATTTTG rACaTtyCTG AAGAACAAC TGTTGATTAT TCTAAAGCCC	780
	ACAATGAACC TTCTTGGATG ACAGAATTAC GTAAAAAAGC TTTGAAATTA ACAGAAACTT	840
30	TAGAAATGCC AAAACCTGAT AAAACAAAAT TAAGAAAATG GGATTTTGAT TCTTTTAAAC	900
	AACACGATGT AAAAGGTGAT GTTTATCAAT CTTTATCACA ATTACCTGAG TCAGTAAGAG	960
	AAATTATTGA CGTAGATCAT TCTAAAAACT TAGTAATTCA ACATAATAAT ACGATTGCGT	1020
35	ACACACAACT TGATGATAAT GCATCGAAAG ATGGCGTTAT CGTTGAAGGT TTAGCAGACG	1080
	CTCTTATGAA CCATAGTGAT TTAGTACAAA AGTACTTTAT GAAAGATGCA GTAACAGTAG	1140
40	ATGAACATCG TATCACAGCG CTACACACGG CATTAGTTAA TGGTGGCGTA TTTGTTTATG	1200
	TTCTTAAAAA TGTAGTTGTA GAACATCCAG TACAATACGT TGTGTTGCAC GACGACGAAA	1260
	ATGCAAGCTT TTATAACCAT GTTATCATCG TTACTIONAAGA	1300

## (2) INFORMATION FOR SEQ ID NO: 461:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3135 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

	GACAGCAAAT TCAAGATACA TTAAATAAAG ATATTGTCAT AAAGCATATT CTTGTTTCGAG	60
	ATAAATCTAA AAAGAGACCG CTAAATATTA GCCAATATCA TTAACTGAA GATGTTAATG	120
5	AAATTTTAAA TGATGATTCA TTAGATATTA TCGTTGAAGT CATGGGAGGA ATTGAACCAA	180
	CTGTAGATTG GTTAAGAACA GCACTTAAAA ATAAAAAACA TGTATTACC GCAAATAAAG	240
	ATTTATTAGC AGTACATCTT AAACTTTtag AAGATTtagC AGAAGAAAAT GGTGTAGCTT	300
10	TAAAGTTTGA AGCGAGTGTA GCAGGTGGTA TTCCGATCGT AAATGCCATA AATAATGGTT	360
	TGAATGCGAA TAATATTTCA AAATTTATGG GAATTTTAAA TGGTACCTCT AATTTTATTT	420
15	TATCTAAAAT GACTAAAGAG CAAACGACAT TTGAGGAAGC ACTTGATGAA GCGAAAAGAC	480
	TTGGTTTTGC TGAAGCGGAT CCAACTGATG ATGTAGAAGG GGTAGATGCA GCGCGTAAAG	540
	TTGTCATTAC ATCATATTTA TCATTTAACC AAGTCATTAA ATTAAACGAC GTTAAACGAA	600
20	GAGGAATTAG TGGCGTAACT TTAACGTATA TTAATGTAGC CGATCAACTG GGTATATAAA	660
	TTAAATTGAT TGGAAGGGA ATATATGAAA ATGGCAAAGT TAATGCATCG GTAGAACCAA	720
	CGTTAATTGA TAAAAAGCAT CAATTAGCAG CTGTAGAGGA TGAATATAAC GCGATTTATG	780
25	TTATTGGTGA TGCCGTTGGT GACACGATGT TTTATGGAAA AGGAGCAGGC AGTTTAGCAA	840
	CAGGTAGTGC CGTTGTCAGT GATTTATTGA ATGTAGCATT ATTCTTTGAA TCAGATTTAC	900
	ACACATTGCC ACCACATTTT GAATTAAAGA CAGATAAAAC ACGGGAAATG ATGGATTCAG	960
30	ATGCAGAAAT TAATATTAAA GAAAAATCCA ATTTCTTTGT AGTAGTGAAT CATGTCAAAG	1020
	GTTCAATTGA AAATTTTGAA AATGAGTTAA AGGCAATATT ACCATTTTAC CGATCATTAA	1080
	GAGTTGCAAA TTACGATAAT CAATCATATG CCGCTGTTAT AGTTGGATTG GAATCATCAC	1140
35	CGGAAGAATT AATCACTAAG CATGGATACG AATTGACAAA GTATACCCAG TAGAAGGAGT	1200
	TTAATTATAA TGAGAAGATG GCAAGGATTA GTAGAAGAGT TTAAAGCACA TTTACCAGTA	1260
40	AATGAAAATA CACCAAAATT AACATTGAAC GAGGGAAATA CACCACTCAT TCATTGTGAA	1320
	AATATGTCTA AAATACTAGG CATAGATTTA TATGTGAGT ATGAAGGTGC CAATCCGACA	1380
	GTTCAITTAAGATCGCGGT ATGGTAATGG CTGTGACAAA AGCAAAAGAG CAAGGTAAGA	1440
45	AAATTGTAAT ATGCGCTTCG ACTGGAATA CATCAGCGTC TGCAGCAGCA TATGCAGCGA	1500
	GAGCAGGTTT AAAAGCTATC GTCGTAATAC CAGAAGGTAA AATTGCATTA GGTAAATTGT	1560
	CGCAAGCAGT AATGTATGGT GCAGAAATCG TTTCTATTGA AGGAACTTT GATGAAGCTT	1620
50	TAGAAATTGT AAAAGAAATT GCAAAAAGTG GCGAAATCGA GCTTGTAAC TCTGTCAATC	1680
	CATTTAGAAT CGAAGGACAA AAGACAGGCT CATTTGAAAT TGTACAACAA TTAGACGGTG	1740
55		



	AAGGCTTTAA AGAATATCAT GAAGCTAAAG GATCACAATT GCCGAAAATG TTTGGCTTCC	1860
	AAGCTGAAGG CGCATCACCA ATTGTTCAAA ATAAAGTCAT TAAAAATCCT GAAACGATTG	1920
5	CAACTGCTAT TCGAATTGGT AATCCTGCTA GTTGGGATAA GGCGACTAAT GCTCTTAAAG	1980
	AATCAAATGG ATTAATAGAT AGTGTTACTG ATGATGAAAT TCTAGAAGCA TATCAGTTAA	2040
	TGACAACTAA AGAAGGTGTC TTTAGTGAAC CAGCGAGTAA TGCTTCTATT GCAGGTTTAA	2100
10	TTAAATTGCA TAGACAAGGT AAATTACCTC AAGGTAAAAA AGTAGTTGCT ATTTTAACTG	2160
	GTAATGGATT AAAAGATCCT GATACTGCTA TTTCACTACT AGATAATCCG ATAAAGCCAT	2220
	TGCCAAATGA TAAAGATAGC ATTATCGATT ATATTAAAGG AGCTTTATAA CATGTCGAAT	2280
15	GTTTGGAGT TAACAATTCC TGCATCAACA GCCAACCTTG GAGTTGGCTT TGATTCTATA	2340
	GGTATGGCTT TAGATAAATT TTTGCATCTG TCTGTAAAGG AAACATCAGG GACAAAATGG	2400
	GAATATATTT TCCATGATGA TGCATCTAAG CAATTGCCTA CTGACGAAAC AAACCTTTATT	2460
	TATCATGTAG CACAACAAGT TGCTTCTAAA TATAGTGTG ACTTGCCCTAA TTTATGTATC	2520
	GAAATGAGAA GTGATATTCC ATTGGCAAGA GGGTTAGGTT CGTCAGCTTC TGCTTTAGTA	2580
25	GGAGCTATAT ATATCGCAA TTATTTTGGT GATATCCAAC TGTCTAAACA TGAGGTATTA	2640
	CAATTAGCGA CTGAAATCGA AGGACATCCT GATAATGTTG CGCCGACCAT TTATGGTGGT	2700
	TTAATCGCTG GATATTATAA TGATGTCTCG AAAGAAACGT CaGtTtGCACA TATCGACATA	2760
30	CCAGACGTGG ATGTGATTGT AACGATACCA ACTTATGAAC TAAAAACAGA AGCATCAAGA	2820
	CGTGCTTTAC CACAAAAATT AACACATAGT GAAGCGGTTA AAAGTAGTGC AATTAGTAAT	2880
	ACAATGATTT tGgCATTAGC ACAGCACAA TATGAATTAG CAGGTAACT CATGCAACAA	2940
35	GATGGCTTTC ATGAACCGTA TCGTCAGCAT TTAATTGCTG AATTTGATGA AGTGAAAACA	3000
	ATTGCTAGTC AACATAATGC CTATGCAACT GTAATTAGTG GTGCTGGACC AACTATTTTA	3060
	ATATTTAGTC GTAAAGAAAA TAGTGGGGAA TTGGTTCGCT CTTTAAATAG TCAGGTAGTA	3120
40	TCATGCCATT CTGAA	3135

## (2) INFORMATION FOR SEQ ID NO: 462:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1209 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 462:

	AGnTCAATAT TTAGATCAAG CCGTTTAAAG TAATTACGAA CAAGTTTATA TCATTCATGG	120
	TAAAGGTACA GGTGCACTTC AAAAAGGTGT ACAACAACAT TTGAAAAAGC ATAAAAGTGT	180
5	TAGTGACTTT AGAGGTGGTA TGCCAAGCGA AGGTGGATTT GGC GTTACCG TTGCAACACT	240
	AAAATAAATT ATAATTTGAT AAATTAAATA GCTGCAGTTA AAATAATGTA AAGCAACAAG	300
	AATACATTTT AACATGTGA TTTGAAATAA GCATAAAAT TGAGCAAATA GAAATACATG	360
10	AAGCATGTGA TCTGATATAA TTTGAACATC ATAATAATAA TTAAGGAGGA TTGGCATTGA	420
	TGGCAATCGT AAAAGTAACA GATGCAGATT TTGATTCAAA AGTAGAATCT GGTGTACAAC	480
	TAGTAGATTT TTGGGCAACA TGGTGTGGTC CATGTAAAT GATCGCTCCG GTATTAGAAG	540
15	AATTAGCAGC TGA CTATGAA GGTAAAGCTG ACATTTTAAA ATTAGATGTT GATGAAAATC	600
	CATCAACTGC AGCTAAATAT GAAGTGATGA GTATTCCAAC ATTAATCGTC TTAAAGACG	660
	GTCAACCACTG TGATAAAGTT GTTGGTTTCC AACCAAAAGA AAACCTTAGCT GAAGTTTTAG	720
20	ATAAACATTT ATAAGTTACA ACCAATGACG ACTGGGGCAT TTCTTTAATG AATTGCTCCA	780
	GTTTTTGTTT GTGTTTTTAA TATAAAAAGT TGAATGATAA GTCATCATAT TGTTTACGAC	840
25	TTGAGAATGG TGGGATTAAT AAATCTATGA ACGTTAAATG ATAATCTAGC ATGCTGATAG	900
	ATTTGTAGCA GTTGGTTTGA TAAAACCATG TTCAATATTA CATGATGTGC ATGAAAAGTC	960
	ATACTCGAAG ATGTTGATTA TTAAGTAGAA TTAGTGGTGA TAAATTTGAA GCACTTTTGT	1020
30	AGCATCATTC ATTTTAAAAT TAGAAGGGGG GATATTTTGG GAAGACTATA AGCAACGAAT	1080
	TAAAAATAAA TTAAATGTCG TACCTATGGA ACCAGGATGC TATTTAATGA AAGATCGTAA	1140
	TGATCAAGTG ATATATGTTG GCAAAGCTAA AAAGCTAAGA AATCGATTGC GATCATATTT	1200
35	TCACGGGTG	1209

## (2) INFORMATION FOR SEQ ID NO: 463:

	(i) SEQUENCE CHARACTERISTICS:
40	(A) LENGTH: 2410 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 463:

	AGTTCAACAC GACGAATTTT ACCTGAGTTT GTTTTTGGTA AGTCGTCAAC GAATTCAATC	60
50	TCTCTCGGAT ATTTATATGG TGCAACTTCA TTTTAAACAA ATTGTTGTAG TTCTTTAACT	120
	AACGTATCAT CACCCGCAGT ATGGTCCTGT AAAATAACGA ATGCTTTAAC AATATTTCTT	180

55

	GCATCTTCAA CTTCAAAAGG CCCAATCGTA TAGCCTGAAC TAATAATAAT GTCATCTCGA	300
	CGTCCTTCAA ACCAGAAATA ACCATCATCA TCTACATGAG CTAAGTCACC AGTGATGTAG	360
5	TATTtACCTG TTTGCGCTTT CGCCGTACGT tCTGGCTCTT tATAATACCC TTTGAAAAGT	420
	GCTGGCAAAT CAAGTGGTAC TGCAATATTC CCTTTCGTAT TAGCAGGTAC GCTATTCCCC	480
	TCATCATCTA CTACAGTGAC CGAACTACCC GGAATGCCTT TACCCATTGA TCCAATCCTC	540
10	TGTGGTGTAT CTTTTAAAAA GCCTATAAGC AAGGTACTTT CAGTCTGGCC ATATCCATCT	600
	CTTACAGTTA AATTAAAGTA TTTCTTGAAT TGTTCAACTA CTTCTCGATT TAGTGGCTCA	660
15	CCTGCAGAAA CGGCACTATG TAAATGCGTT AAGTCATAAT CATTTAAGTT CTGTAATTTA	720
	GCCATCATAC GATATTCTGT CGGTGTACAA CATAAAACAT TAATTTGATA TTTTGAAGC	780
	AATTCTAAGT ATGTTTCAGG ACTGAACCTT CCATTAAATA CAAAAGCAGT TGCACCTGAA	840
20	CCTAATACAG ATAAGAAAGG ACTCCATACC CATTTTGGCC AACCTGGTgc TGCTGTTGCC	900
	CAAACtaAGT CATCTTCATT aATACATaAC CAATGTTTTG GTGCCATTtG TaAATGTGcA	960
	AATCCCCaTC CATGACAATG TGTAACGGCT TTAGGATTGC CAGTTGTACC AGATGTATAT	1020
25	GACAGAATCG CCATATCATC ACGCGTCGTA TCTGCCATTT CTAGTTTGTT ACTTGCGTTT	1080
	TCTTTTTcAG CTTCAAGTGA AATCCATCCA TCTTTTTGAC CGGCAATAAC AAATTTAGTT	1140
	AACGCATCAT ATTCTTTAAT TTTTTCAAAT TCAACTGTGA ATGGCTCTAG TGCAATAACT	1200
30	GCATTAATTT CACCATGTGT GATACGGTAT TGTAATCTT TAGTTCTTAG CATTTcAGAA	1260
	CATGGAATGA TTGCAACACC TAATTTTAAA GCAGCAATAT ATAATTCATA CGTCGCAATA	1320
	GATCGTGGCA TCATAATGAG TACTTTATCG CCTTTAGATA AACCGTGCGA TGCTAAAACA	1380
35	TTACCTACTT TATTAGACTG TTCAATGAGC TGTtGGTAAG TGACTGATAT ATCTTCGCCT	1440
	TCAGTATTAT GATATAAAAT TGCCTTTTTA TCTGGTATGT GGCTATATTT TTCGATTtCC	1500
40	GAAATAATGT TATATTTTTC AGGCGCGAAT AGAGCTGACT TTTGCATAAC TAACTTCCTT	1560
	TCATACATCC ACTTTTCCTG TGATGAACAT TGTAATTTTA TAAATGAATT ATATACATCA	1620
	TACGCCTATC TTTACAGAAT TTTCAATTAA ATAGGGTTAA ATACCAAAGT CCTCGACACT	1680
45	ACACTTTGAC ATGACGTAGC ATTCAAGGAC TTTCAAATGA TTGAGGGTTG ATATCTCGGG	1740
	CTAGACCATA TCAGCTAATT CAATACGAAT ATTGTATGAT AATTCACGAT TAATTATTTT	1800
	TACATCTGCA CCTTTCGAAG TGCCACGATG CTTGTGTGTA TGCTTGTA CT CAGCTGAATT	1860
50	TTGCCAATGA TAAATGCTT GCCTATTTTC CCACAGCGTA ATAATGATAT AGTGTCTACC	1920
	AGCTGTTCTA GGTCTTAAAA ACCTTAATGC TTTAAATCCA TCAACGTTTT TTAAATGCTT	1980
55		

ATTTAACACA CATAATGAAT CATTTGATAA ATCATTTTATC GCTTCTAGCA CATCGTAATA 2100  
 TGCAGTGTCA TTATTTTTTT GTATTGTGAG ACAATCATCC AGTTCTTCTA TTACATAACT 2160  
 5 TCTATATTCA TCATAAATTT TCATAATAAA TGCCTTCATT TCATTTATAT TTTTGGTCAT 2220  
 ATTACTkTAT ATCTATTACT AatkCATTCC CGTATTTTATT AATTACAATC ATAGTTTGGC 2280  
 TyCTTTTTTAA AAGATAAGAC TTTGTAAAAA GTATTAATAT TTCATGCAAA TGGGGGACAG 2340  
 10 GAGTCGCCCA CTATTTTGT GTCTTCAATT TCATGATCAT TATTTAACAT TAGTCATGAA 2400  
 AATAGCCGAC 2410

15 (2) INFORMATION FOR SEQ ID NO: 464:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 590 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 464:  
 25 TTTATTAATT GTAAAAAATT GAGTAAATTA TCTTTACATT CTAAATTAGT CTTAACTACA 60  
 ACTAGTATCC TAATAATTAT AGGAGCTATT ACATTCTTTT TATTAGAACA GTTTAATACT 120  
 ATGCAACATA TGGGACTAGT TGAAAAAATC GGAAATTCTT TTTTCCAATC AGTAACAACA 180  
 30 CGAACAGCGG GTTTTAAACAG TATAGATATA GCAAGCATTa rCAAATCTAC CGCATtAaTG 240  
 TTAATGCTAC TTATGTTTAT TGGTGGTGCC CCTCTCAGTG CAGCTGGAGG AATTAAAATA 300  
 ACTACTTTTG CAGTTGCGTT TATTTTGTGA CTAAATTATA kACGTAAAGA AAATAATGTT 360  
 35 TCAGTATTCA ATAAAGAAAT ATCTGACAAA CATATAAAAC TATCTATTGT TACCATTAAAT 420  
 ATCTCATTTC TATTTATCAG CATCATTACT TTTATATTAT CGATAATTAA TCCGAACATA 480  
 40 TCATTAATCA AGTTATTATT CGAAGTGGTT TCTGCATTTCG GAACAGTAGG GTTAAGTATG 540  
 AACCTTACCA CAGAATATCA TGGTATTACT AAAATAATTA TTATATTCGT 590

(2) INFORMATION FOR SEQ ID NO: 465:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 905 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 465:

55

TATTCGATTT GAnTCTTTTA AAATTATGTT TAAAAACATG TCTAATGATT CCGCATCATT 120  
 TTTTGCACCA TCAATAAGCG TTTCAGcAAA CCCCTTAATT GAAGTAATAG GTGTTTTTTAA 180  
 5 TTCATGTGAA ACATTTGCTA CAAATTCACG TCTTAGATTT TCAAGTTGTT TCAGATTTGT 240  
 TATATCATGC ATCACAaACTA AAATCCcTG CAAACTTTTT TGAGACCTAG TTAAAATCGG 300  
 AACGCATGAA ATATCAAAGT ACTTGGCATG GACTTGTTTT ATTGCAACTT CCAATTGTTC 360  
 10 ATAAATAGGT TTTTCAACTT TAAAACTTTC TAAAATTAAT TGCTCAATTT CAGTATTAAAC 420  
 ATAGCCGTGA TAGCCTACTT GTTCAATATT ATGCGAGATG TTGAAGTGT CATAATACGC 480  
 TTTATTTGCA ACAACGATTT TTCCATTTTCG ATCTATCATT AAAATAGCAC TTGGAATATT 540  
 15 TTCAATCGTT GTTTTTTAAAC GGTGGAATTG AATTTTTTGC TCATTATTAA GCTTTTGAAG 600  
 GCGTCGTGCT AAATCATTGG TAGACACAAA AAGCGCTTTA GTTTCTACAA CATTACTTTC 660  
 AGGTACACGT ATGTGATAAT AACCATTTGC CAACAATTGT GTTGCATAAG TAACTTCTTG 720  
 20 AATGGGACGG ATTAATGTAC GCTTAAaACT ACGGCTTGCA AAATACAGAC AAATGAGTAC 780  
 AACTAAACAT GTCAAAATAA GATATTTCCA CAACGTCCAA TGCATTCTG TAATATCGTT 840  
 25 ATTGTAACCT TTAATCCATA CATGATAACC GTTAACCTTC TTATtAAAAA TAAAAACGTC 900  
 CCTTT 905

## (2) INFORMATION FOR SEQ ID NO: 466:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1016 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 466:

TTTGGTTAGC CCATAAAAGA AAAAAACAAG TAGTCATTTT TAAACAACAT ATCAAGTCTA 60  
 40 CCCAAGAAAT ACGTTTTGAC AAAGCGAAAG TGCTTGAACA CAAAGATGAA ATAGCAAATT 120  
 TTATTTCTTT CGAACCACAA AGTTTGAAT TTTATTATTT TACAGAATCG GAATTTTCAG 180  
 AAGAACAATT AAATGAAGTT TCGCCAATTA GAATTAAATT CAATGTTATA AGACACACAA 240  
 45 AAGATTTGaT AAAGCATATG CCGAATATAT TTTTGGcTAG ACTTATTTCa GAAGATAATG 300  
 ATAAAAAGAC ATATATGTTT TATAAACGCA AAGTATTAAC CGATAACTTT TTAGATAAAT 360  
 50 ATATGCAGAA ATTTTCACCG GCAACATACA CAATAATATT TGTAATGTC TTAATATGGT 420  
 TATGTATGAT TTTATATTTA AATAATTTTT CGGATGTAAA ATTATTAGAT GTTGGCGGGT 480

55

ATTTTAGTTT TGAACATATA CTTATGAATA TGCTTTCATT ATTTATTTTT GGTAAGATAG 600  
 TCGAAGCAAT TATTGGTTCA TGGCGGATGT TAACTGTATA CTTTATTGCA GGGTTGTTTG 660  
 5 GAAACTTTGT ATCACTATCA TTAAATACGA CTACAATTTC AGTTGGGGCT AGTGGTGCTA 720  
 TATTGGTCT GATTGGATCA ATTTTTCGCA TGATGTATGT TTCAAAAACA TTAAACAAAA 780  
 AAATGTTAGG ACAGTTATTA ATTGCATTAG TGATATTAGT TGGTGTCTCT CTGTTTATGT 840  
 10 CAAATATAAA TATTGTGGCG CATATTGGAG GATTCATTGG TGGTTTATTA ATAACTTTAA 900  
 TTGGCTATTA CTATAAAGTG AATCGtAATA TTTTTCGaT TTaCtAATTG GTATGCTtGt 960  
 tATATTTAwT GCACyTCmAA TTagAtTTTT ACmATTAAAG AAGATAATAw TTATAA 1016  
 15

(2) INFORMATION FOR SEQ ID NO: 467:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 406 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 467:

AACTTTAAAT TTAGACATCT TTAAAACCTC TCTTAAACCA TGCCTATATC TCAAGATGAT 60  
 ATTTCAAATG AACAACTACTA TTGCTTGAGA CCATTAATGA ATGATCATAA ATATTTCTTT 120  
 30 CTATAAAATT AGCTTTCCAA TAACTGTGTT GTTGCAATAT ATCATTCACA AGTACACCAT 180  
 TTTCGGAAGT ATGATTATCT TTATCTATAC TTAAACAAT TTGTTTAGTT TTAGCATGGC 240  
 TAAATTGTG AAGACCTTA CACGATAAAC GTATAGCGTC TGAATTCTCA TTAAACAATG 300  
 35 CGGCTGGgCA AACAAATGAC ACATTGTACT TCATGTTTGA ACTTCGTTAC AATCaTCGTG 360  
 kCatTTTGat AAATAACAAT CCCTCGTAAT kGATTAAGTA TATTAT 406

(2) INFORMATION FOR SEQ ID NO: 468:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1378 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 468:

50 AGATAATACA CTTGAAGTTG GAATGGTTTG TGACGGTTAT TTAATGCGAA TTGAAAACCTT 60  
 AACACCATCA AATTTCTTCA ACTCAGCAAG TGAAGATACG ATTACTAAAA TTAAATTAAA 120

55

AGGTACAGCG TTAAACTAA GAGAAGCCAT CAATTATGAT GAAATGGTTA TTGTAGATAG 240  
 TATGACGTAG TTCCTAATTA TGCKAAAAGG GATTGATGAA AAAGTGAAGG GCTTTTCATC 300  
 5 AATCCCTTTT ATTTTAGGGG AATTGAATAG ATAGTTTAA ACTATACGAA TTATTAATAT 360  
 TTGAGATTTA ATTGAAATAA GTTTTAAAAA TTGGAGGAGA TAGATTAAAGC GAAGTCATTT 420  
 AAAGGTGAAG TTAAGTGTAT TCACAAAAAn TAGCCACACT CATATGACAT CGGATGAGTG 480  
 10 TGGCTTAAGG ATCTATGGGG GGAGGAAnCC ATAGATGTTT ACTTTGATAG GCCAGATTAA 540  
 ATATCAAAGT ATGCGATTAT TTATAGCTTG ATGCAAAAGT GGTATGCCTA TTAAGTTA 600  
 CTGCACATAG CTTTTAATAT TCCGTTCAAA GGAAAGGGGC ATACAATTGA ACAATCTGTA 660  
 15 ATAGTACTTT TAACCAGCTA TGCTAAAAGT CTAGTAGGGA GAACAGTTGT CCAATCACAT 720  
 AAGAACCTCT AACTTCGTTA GTACGATTAA GAAAAGCTTT TTAGTTAGTA TGTAATACAA 780  
 20 TTTATTGACG CGCGTGAATC TCTTTTATAA GAGTGTGTAG GGAATGGCGT TGTATAAATT 840  
 GTATTAGAAG AACTTCTAAC GCATCTCTGT GGTAAAAAGA GATGAAGGGA ACGACAGTTT 900  
 aATTAAAACT GCATAAGAAC TTCTAGCTTT TCTCTCTCGT TCAAAGAGAA GCAGcTGTTc 960  
 25 GCAGTTTAAT CAAAACCACA TAAAGCTTTT AACTTTACTC TTTGATTTAA AGAGTGATAA 1020  
 ATGTTTACAG TTAAATTAAA ACTGCATAAG AACTTCTAGC TTTTCTCTT CGTTCAAgAG 1080  
 AAGCAGCTGT TCGCAgTTTA ATCAAAACCA CATAAAGCTT TTAACTTTAC TCTTTGATTT 1140  
 30 AAAGAGTGAC AAATGTTTAC AGTTTAATTA AAAGTGCATA AGAACTTCTA GCTTTTCTCT 1200  
 TTCGTTCAAA GAGAAGTTCT AATACCACCA TATCGTGCGA TCGGGAACGG TATATATATT 1260  
 AATAGGAGGG TAATATATAT TTAACGCACG ATATGGGACT ATTAGCCTTC GACTTTGTTA 1320  
 35 TGTTGATGTG TGGCCTAAAA TATTGGAGAT ACCAATATTT TAGGTTGCAT CAACATCA 1378

(2) INFORMATION FOR SEQ ID NO: 469:

(i) SEQUENCE CHARACTERISTICS:  
 40 (A) LENGTH: 4171 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 469:

TCCCAACCAA TAATCGTGGC AAAAATACGG ATATTGGTAT GGCTTAACAA ATTGCAAATA 60  
 50 TCGTTTAATC ATACATCCCC CCTAATCTAT TGCCCTATCC TATTCaTAAG CATAAAAATG 120  
 AATAGAGGTT GGATACATAA TTTGTAGATG TAAATCTTC TTACAATTTA CATTTTTTAA 180

55

TTATATTCTA TTCAATTTAA TCTATGGATA CTGTGTCCCC ACACGACAGC AAAAGTTATC 300  
 ATACTTCTTT ACATCACTAA GTCAATATAA ATGATTTAAT CAGTATTTAC ACTTTATTTG 360  
 5 CTTAATACTG TCTAATTTTT TTGTAACGTT CTTTCCAAAC TTTGATAAAA TCTGGCGCGA 420  
 ATGGGCCCTT CTTCTGTCT ATCCATTGTT GAAGAATGTC CACGTTGCGT CTTAAAATAA 480  
 TATCAATATC ATGCGGATAA TTCATTTGAT TCATATGTTG CTCATATTCA TCTTCATCTA 540  
 10 ATAAATGATA CTTTCCGTTT GGATATACTT TAATATCTAA ATCATAGTCT ATATATTTTA 600  
 ATGCCTCTTC ATCACAAACA AATGGTGATG ACAAATTGCA ATAGTAATAA ATTCCATCTT 660  
 CTCTAAACAT GCAGATAACA TTaAACCAAT ATTCTGAGTG AAAGTaAACA ATTGCCGGTT 720  
 15 CACGTGTTAT CCAAGTTCTT CCGTCACTTT CAGTCACTAA CGTATGATCA TTTCCACCAA 780  
 TGACAACATG ATCAGTACCC TTTAATATTG TTGTTTCAGA CCAAACGCGA TGAATCTTAC 840  
 20 CATCATGTTT ATAACCTGTA ATTTAATGT TTTCCCTTC TTTAGGTATG GATTCTCTGA 900  
 CCATACTCCA CACCACCTTC TGTAATTTA ACCATTATAA ATTATAGCAT ATTTCAGAAA 960  
 TAGTATTATA TAAATACATA TTTTACGAA ATAAGATTTT ACTACTTAAT AATTAAACTC 1020  
 25 GGTAATATTG CTAAGTACTA CAACAGAGAT TTACATGTCC CATTTAAAGT ATATAAAATC 1080  
 ATCACTTTTA TATATCAACA CTTTAACTTT TTGACATTGT TATTCTATGA GATTTAAAGA 1140  
 TATCATTTAT ACTTTTTTAAA ATTAATGTCA CTATGTTTTT CGATAATATT ACCAATCATC 1200  
 30 GAATGTTACC CATTTATAAA TTGATAAATs TTTGACATAG GTACAGGGAA TGTATATTGA 1260  
 TCTCGATCAC TTAAATCAAA CCAAATCATG TCATCTGGTA ATGTTTCAAT GTTAATTGCT 1320  
 CCTGAAACGG CGTATACTTT AATCTCCAT GTTAAATGAG TAAATTGATG CTTCAACTCA 1380  
 35 AAAATAGGTG TTTCTACTGG TTGAATGTCA TGACCGATTT TTTCAGTCAT TTTACGTCTA 1440  
 GCATGCTCAC TTTCAACAT AGGAAATTGC CACATACCAT GCAATAATTT TTCGCTACGC 1500  
 TTTTGCAACA GATATTGACC TTGATTATTT CTAATTAAAA AGACGGATTG CTCAATTACT 1560  
 40 TTTTACTTA CATTTTTAGA TTTAACAGGT AACTTTTCAA ATGTACCTTT ATCAAATGCC 1620  
 TCACAGTTTT CTTGAACCTG ACAAATAAG CATAATGGAT TTTTGGTGT ACAAATTAAC 1680  
 45 GCCCCTAATT CCATCATAGC TTGATTAAAC GTTCCAGCTT CTGTAGTAAC ATACGGTAAT 1740  
 AATTCTTGTT CGTACGATTT CCTCGTCGAT TGTAATTTAA TATCTCGATA GTCATCATTC 1800  
 AATCTAGACC ATACTCGAAA AACATTTCCG TCTACAGTTG CTAGTGGTAC ATTATATGCA 1860  
 50 ATGCTCATTA CTGCAGCTTG TGTGTATGGG CCAACACCTT TTAACGCTTT AAATTGATCA 1920  
 GGATCTTTGG GAACTAAGCC TTCATATTTA TCATGAACCT CTTAATCGC CGTATGAAAA 1980

55



GCTTGaCTCA AAACCTCCAC AGTTGGAAAT CGTTCAACAA AACGATGATA ATAGTCAATA 2100  
 ACTGTTTAA CTGTGTCTG TTGTAACATG ACCTCACTTA ACCAAATATA GTACGGATTG 2160  
 5 GTCGTTTGTC GCCATGGCAT TTCTCTTTGA TTTTCATCAA ACCAGTGTAT CAAATTTTCT 2220  
 TTAAACTAG ACTGCTGATA CATTTATAAA ACCCTTTCCT CACCAAAATT AATTGTCTTT 2280  
 ACTCATAATG TTTTATTGT ACATTAAAT CATGGTTAGT ATGTAAGTTA ATTTAGTTAT 2340  
 10 TTGCGAAATT GGATTATAAT AGTATATATA ATATTATGAA ATGAGTGAAC TGATATGGAC 2400  
 ACTGCAACAC ATATCGCAAT TGGGGTGGGC CTTACAGCAC TTGCAACTCA AGATCCAGCA 2460  
 ATGGCTTCTA CGTTTGGTGC AACAGCTACA ACCCTTATCG TTGGTTCATT AATTCCTGAT 2520  
 15 GGGGATACTG TTCTTAAATT AAAGGACAAT GCAACATATA TTTCGCATCA TAGAGGTATC 2580  
 ACGCATTCCA TCCCTTTCAC AATACTATGG CCAATTTTAA TTACATTTT AATATTCACG 2640  
 20 TTCTTTAGTG GAACCAACCC ATTCATGTA TGGATGTGGG CTCAGCTCGC AGTATTTTAA 2700  
 CATGTCTTTG TAGATATATT CAATTCCTAT GGTACACAAG CGCTTAGACC TATCACAAC 2760  
 AAATGGATTC AATTAAGTGT GATTAACACA TTTGACCCTA TTATTTTCAC AGTTCTTTGT 2820  
 25 ATTGGTATTG TATTATGGGT TATAGGCTTG CATCCATTG CAGTCTTCTT TCCTATAATC 2880  
 GCTTTACTAA TCATTTATTA CATGATTCGT TTTAAATGA GAGCCGTAAT TAAGCAACAA 2940  
 GCTTTAAAAG CAATTCAACA AGAGCATCAC CCTGTTAAAG TATTTGTTGC GCCAACAATA 3000  
 30 AAATTTATGG AATGGCGTGT CGCGATACAA ACTGATGCAC ATGACTATGT TGGAAAAGCA 3060  
 TATGGTAGGA ATGTGGTGTT TAGTGATAAA GTGGAACGTC AAACATTATC AACAGACTCC 3120  
 ATTTTATGGA AAGTCAAAGG TAATAAAGAT ATACGTACTT TTTTAACTT TTCATCAATC 3180  
 35 TATCGTTGGC AAACAACAAC GTTAGCAGAT GGTCTACTG AAATTCGTTT GATTGATTG 3240  
 CGTTATTTAA AAAATGATCA TTATTCATTT GTGGCAATTG CACATGTAAC AAACGATAAT 3300  
 GTCATAGACC ACTCTTATAT TGGCTGGGTA TTTACAGAAG ATAAGTTACA ACGTAACTG 3360  
 40 TATGCTAAAT AATTTCAAGT TATTATTCAC TAAAGTTAAT CTATAAAAAA TGAACAACCG 3420  
 GGCAGAATGA AAATCAAAAC GATTTTACT CTGTCGGTT TTTAATGTA AAATATGAA 3480  
 45 TGCTTTTACA AAATCTAAAA TTTATATTGT TGCTAACAAA CTACCTTTAA TGAATCGAAA 3540  
 TATCAAAATC AGTATAGGAA AACAATATCT AGATGATATT CTAATTGTTT CTGATTCTCA 3600  
 CAGATTAATT TACACAACAG GTCAGCTAAA CATCATGAAG AAGTATCCGC CTCGTCTGTA 3660  
 50 CTATCATTTG AAACATCCTG TTGATTATCa GTTTGTGTCA CTTTAGATGT TTTATTaTAA 3720  
 ATTGCGTGTG TCGTATACCT TGCTAATACG AAATTTACTG CAGCTATTAA ACATAAGAAT 3780

55

TGTGCTATAC CATTAAACAAT GTAATACATT GGATTTAGCA TTAGGATGTG ATTGATAAAAT 3900  
 ACATGATTTG GATTTGGTAT GAAAATAATT GGTAACAATA AGAAACACAA TACACAAACC 3960  
 5 CCATAAAATA TGATATTTAT TTTTTCAGWT AACAGTCGAA TAAGACCAAA AGTAACGGAT 4020  
 ATTAATCCTA CAAAAATAGT TGCCATCACA ATAAATAGA ATAGCGCTAT ATATGATGTT 4080  
 TCGAAGTTTA CTGGTTTAAC CAATGCACTA ATCATCGTCA AAATGACTAG CATAATAAAA 4140  
 10 CTTAAAATAG ACATAATAAC TACTGGCGTC G 4171

(2) INFORMATION FOR SEQ ID NO: 470:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 9821 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 470:

TGGTTGAAGT AGCAGTTAAT TCAAAATCTG CAACAGTTTC AGCAGAAATaG GGGCTTTCAA 60  
 25 AATAAtCAAA GGAGAATAAT TTATGACTAA AACTTTAAAG GTTTATAAAG GAGACGACGT 120  
 CGTAGCTTCT GAACAAGGTG AAGGCAAAGT GTCAGTAACT TTATCTAATT TAGAAGCGGA 180  
 TACAACTTAT CCAAAGGTA CTTACCAAGT GGCATGGGAA GAAAATGGTA AAGAATCTAG 240  
 30 TAAAGTTGAT GTACCTCAAT TCAAAACCAA TCCAATCTA GTCTCAGGCG TATCATTTAC 300  
 ACCAGAAACT AAATCAATTA TGGTAAATAC CGATGACAAT GTTGAGCCAA ACATTGCACC 360  
 AAGCACAGCA ACGAATAAAA TATTGAAATA TACAAGTGAA CATCCAGAAT TTGTTACTGT 420  
 35 AGATGAAAAT ACAGGAGCAA TTCACGGTGT AGCTGAAGGT ACTTCAGTAA TCACTGCTAC 480  
 GTCTACTGAT GGAAGCGATA AGTCAGGACA AATTTTCAGTG ACAGTAACAA ACGGATAGGG 540  
 ATTTAAGGCG CAGTATATCT GCGTCTTTTT TATTTGAATA AAAGGAGCTA ATACAATGAT 600  
 40 TAAATTTGAA ATTAAAGATC GTAAAACAGG AAAAACAGAG AGCTATACAA AAGAAGATGT 660  
 AACAAATGGC GAACAGAAAA ATGCTATGAG TATTTAGAAT TAGTAAATCA AGAGAATAAA 720  
 AAAGAAGCAC CTAACGCAAC AAAAATGAGA CAAAAGAGC GACAGTTATT AGTAGATTTA 780  
 45 TTTAAAGATG AAGGATTGAC TGAAGAAGAT GTTCTGAACA AGATGAGTAC TAAAACCTAT 840  
 ACAAAGCCT TACAAGATAT ATTCGAGAA ATCAATGGTG AAGATGAAGA AGATTCAGAA 900  
 50 ACTGAACCAG AAGAGATGGG AAAGACAGAA GAACAATCTC AATAAAAGAC ATTTTATCGA 960  
 ACATTAAGAA AATACAACGT TTCTGTATGG AGCAGTATGG GTGGACATTA ACTGAAGTCA 1020

55

	AAGAAAAACA AAGTGAACAA AAAGTCATTA CAGGTACGGA TTTAAGAAAA CTTTTTGGA	1140
	GCTAGAAAGG AGGTTAATAT GAATGAAAAA GTAGAAGGCA TGACCTTGGA GCTGAAATTA	1200
5	GACCATTTAG GTGTCCAAGA AGGCATGAAA GGTTTAAAGC GACAATTAGG TGTTGTTAAT	1260
	AGTGAAATGA AAGCTAATCT GTCAGCATTT GATAAGTCTG AAAAATCAAT GGAAAAATAT	1320
	CAGGCGAGAA TTAAGGGGTT AAATGATAGG CTTAAAGTTC AAAAAAGAT GTATTCTCAA	1380
10	GTAGAAGATG AGCTTAAACA AGTTAACGCT AATTACCAAA AAGCTAAATC CAGTGTAATA	1440
	GATGTTGAGA AAGCATATTT AAAGTTAGTA GAAGCCAATA AAAAAGAAAA ATTAGCTCTT	1500
15	GATAAATCTA AAGAAGCCTT AAAATCATCG AATACAGAAC TTAAAAAGC TGAAAATCAA	1560
	TATAAACGTA CAAATCAACG TAAACAAGAT GCGTATCAAA AACTTAAACA GTTGAGAGAT	1620
	GCAGAACAAA AGCTTAAGAA TAGTAACCAA GCTACTACTG CACAATAAA AAGAGCAAGT	1680
20	GACGCTACA GAAGCAGTCC GCTAAGCATA AAGCACTTGT TGAACAATAT AAACAAGAAG	1740
	GCAATCAAGT TCAAAAATA AAAGTGCAA ATGACAATCT TTCAAAATCA AATGATAAAA	1800
	TTGAAAGTTC TTACGCTAAA ACTAATACTA AATTAAAGCA AACAGAAAA GAATTTAATG	1860
25	ATTTAAACAA TACTATTAAG AATCATAGCG CTAATGTCGC AAAAGCTGAA ACAGCTGTTA	1920
	ATAAAGAAAA AGCTGCTTTA AATAATTTGG AGCGTTCAAT AGATAAAGCT TCATCCGAAA	1980
	TGAAGACTTT TAACAAAGAA CAAATGATAG CTCAAAGTCA TTTCGGTAAA CTGCAAGTC	2040
30	AAGCGGATGT CATGTCAAAG AAATTTAGTT CTATTGGAGA CAAATGACT TCCCTGGGAC	2100
	GTACAATGAC GATGGGCGTA TCTACACCGA TTACTTTAGG TTTAGGTGCA GCATTAAAA	2160
	CGAGTGCAGA CTTTGAAGGG CAAATGTCTC GAGTTGGAGC GATTGCACAA GCAAGCAGTA	2220
35	AAGACTTAAA AAGCATGTCT AATCAAGCGG TTGACTTAGG AGCTAAAACA AGTAAAAGTG	2280
	CTAACGaAGT TGCTAAAGGT ATGGAAGAAT TGGCAGCTTT AGGCTTTAAT GCCAAACAA	2340
40	CAATGGAGGC TATGCCAGGT GTTATCAGCG CAGcAGaAGC AAGTGGTGCA GAAATGGCTA	2400
	CAACTGCAAC TGTAATGGCT TCAGCGATTA ACTCTTTCGG TTTAAAAGCA TCTGATGCAA	2460
	ATCATGTTGC TGATTTACTT GCGAGATCAG CAAATGATAG TGCTGCAGAT ATTCAATATA	2520
45	TGGGAGATGC ATTTAAATAT GCAGGTACTC CAGCAAAAGC ATTAGGAGTT TCAATAGAGG	2580
	ACACTTCTGC AGCAATTGAA GTTTTATCTA ACTCAGGTTT AGAGGGGTCT CAAGCAGGTA	2640
	CTGCATTAAG AGCTTCGTTT ATTAGGCTAG CTAATCCAAG TAAAAGTACA GCTAAGGAAA	2700
50	TGAAAAAATT AGGTATTCAT TTGTCTGATG CTAAAGGTGA GTTTGTTGGA ATGGGCGAAT	2760
	TGATTAGACA GTTCCAAGAT AACATGAAAG GCATGACGAG AGAACAAAA TTAGCAACAG	2820

55

EP 0 786 519 A2

	CAGATAAAAT TAATAGCTAT AGCAAATCAT TGAAGAACTC TAATGGTGAA AGTAAAAAAG	2940
	CAGCTGATTT GATGAAAGAT AACCTCAAAG GTGCTCTGGA ACAATTAGGT GCGCTTTTG	3000
5	AATCGTTAGC AATTGAAGTT GGTAAAGATT TAACGCCTAT GATTAGAGCA GGTGCGGAAG	3060
	GATTAACAAA ATTAGTTGAT GGATTTACAC ATCTTCCTGG TTGGGTTAGA AAGGCTTCGG	3120
	TAGGCTTAGC AATTTTTGGT GCATCTATTG GTCCTGCTGT TCTTGCTGGT GGCTTATTAA	3180
10	TACGTGCAGT TGGGAGCGCG GCTAAAGGCT ATGCATCATT AAATAGACGC ATTGCTGAAA	3240
	ATACAATTCT TTCTAATACC AATTCAAAG CAATGAAATC TTTAGGTCTT CAAACATTAT	3300
15	TTCTTGGTTC TACAACAGGA AAAACGTCAA AAGGCTTTAA AGGATTAGCC GGAGCTATGT	3360
	TGTTTAATTT AAAACCTATA AATGTTTGA AAAATTCTGC AAAGCTAGCA ATTTTACCGT	3420
	TCAAACTTTT GAAAAACGGT TTAGGATTAG CCGCAAAATC CTTATTTGCA GTAAGTGGAG	3480
20	GCGCAAGATT TGCTGGTGTA GCCTTAAAGT TTTTAACAGG ACCTATAGGT GCTACAATAA	3540
	CTGCTATTAC AATTGCATAT AAAGTTTTTA AAACCGCATA TGATCGTGTG GAATGGTTCA	3600
	GAAACGGTAT TAACGGTTTA GGAGAACTA TAAAGTTTTT TGGTGGCAAA ATTATTGGCG	3660
25	GTGCTGTTAG GAAGCTAGGA GAGTTTAAAA ATTATCTTGG AAGTATAGGC AAAAGCTTCA	3720
	AAGAAAAGTT TTCAAAGGAT ATGAAAGATG GTTATAAATC TTTGAGTGAC GATGACCTTC	3780
	TGAAAGTAGG AGTCAACAAG TTTAAAGGAT TTATGCAAAC CATGGGCACA GCTTCTAAAA	3840
30	AAGCATCTGA TACTGTAAAA GTGTTGGGGA AAGGTGTTTC AAAAGAAACA GAAAAAGCTT	3900
	TAGAAAAATA CGTACACTAT TCTGAAGAGA ACAACAGAAT CATGGAAAAA GTACGTTTAA	3960
	ACTCGGGTCA AATAACAGAA GACAAAGCAA AAAAATTTT GAAAATTGAA GCGGATTTAT	4020
35	CTAATAACCT TATAGCTGAA ATAGAAAAAA GAAATAAAAA GGAATCGAA AAAACTCAAG	4080
	AACTTATTGA TAAGTATAGT GCATTCGATG AACAAGAAAA GCAAAACATT TTAAGTAGAA	4140
40	CTAAAGAAAA AAATGACTTG CGAATTAAAA AAGAGCAAGA ACTCAATCAG AAAATCAAAG	4200
	AATTGAAAGA AAAAGCTTTA AGTGATGGTC AGATTTCAGA AAATGAAAGA AAAGAAATTG	4260
	AAAAGCTTGA AAATCAAAGA CGTGACATCA CTGTTAAAGA ATTGAGTAAG ACTGAAAAAG	4320
45	AGCAAGAGCG TATTTTAGTA AGAATGCAAA GAAACAGAAA TGCTTATTCA ATAGACGAAG	4380
	CGAGCAAAGC AATTAAAGAA GCAGAAAAAG CAAGAAAAGC AAGAAAAAAA GAAGTGGATA	4440
	AGCAGTATGA AGATGATGTC ATTGCTATAA AAAATAACGT CAACCTTTCT AAGTCTGAAA	4500
50	AAGATAAATT GTTAGCTATT GCTGATCAAA GACATAAGGA TGAAGTAAGA AAGGCAAAAT	4560
	CTAAAAAGA TGCTGTAGTA GACGTTGTTA AAAAGCAAAA TAAAGATATT GATAAAGAAA	4620

55

## EP 0 786 519 A2

	GTTGGTGGTC	TAACCTTTAGA	GAAGACCAAA	AGAAGAAAAG	TGATAAATAC	GCTAAAGAAC	4740
	AAGAAGAAAC	AGCTCGTAGA	AACAGAGAAA	ATATAAAGAA	ATGGTTTGGA	AATGCTTGGG	4800
5	ACGGCGTAAA	AACTAAAACT	GGTGAAGCCT	TTAGTAAAAT	GGGCAGAAAT	GCTAATCATT	4860
	TTGGCGGCGA	AATGAAAAAA	ATGTGGAGTG	GAATCAAAGG	AATTCCAAGC	AAATTAAGTT	4920
	CAAGTTGGAG	CTCAGCCAAA	AGTTCTGTAG	GATATCACAC	TAAGGCTATA	GCTAATAGTA	4980
10	CTGGTAAATG	GTTTGAAAAA	GCTTGGCAAT	CTGTTAAATC	GACTACAGGA	AGTATTTACA	5040
	ATCAAATAA	GCAAAAGTAT	TCAGATGCCT	CAGATAAAGC	TTGGGCGCAT	TCAAAATCTA	5100
15	TTTGGAGAGG	CACATCAAAA	TGGTTTAGCA	ACGCATATAA	AAGTGCAAAG	GGTTGGCTAA	5160
	TAGATATGGC	TAATAAATCG	CGCTCGAAAT	GGGATAATAT	TTCTAGTACA	GCATGGTCGA	5220
	ATGCAAAATC	CGTTTGAAAA	GGAACATCGA	AATGGTTTTAG	TAACATACAT	AAATCTTTAA	5280
20	AAGGTTGGAC	TGGGGATATG	TATTCAAGAG	CCCACGATCG	TTTTGATGCA	ATTTCAAGTT	5340
	CGGCATGGTC	TAACGCTAAA	TCAGTATTTA	ATGGTTTTAG	AAAATGGCTA	TCAAAAACAT	5400
	ATGATTGGAT	TAGAGATATT	GGTAAAGACA	TGGGAAGAGC	TGCGGCTGAT	TTAGGTAAAA	5460
25	ATGTTGCTAA	TAAAGCTATT	GGCGGTTTGA	ATAGCATGAT	TGGCGGTATT	AATAAAATAT	5520
	CTAAAGCCAT	TACTGATAAA	AATCTCATCA	AGCCAATACC	TACATTGTCT	ACTGGTACTT	5580
	TAGCAGGAAA	GGGTGTAGCT	ACCGATAATT	CAGGAGCATT	AACGCAACCG	ACATTTGCTG	5640
30	TATTAAATGA	TAGAGGTTCT	GGAAACGCCC	CAGGTGGTGG	AGTTCAAGAA	ATAATTCACA	5700
	GGGCTGACGG	AACATTCCAT	GCACCCCAAG	GACGAGATGT	GGTTGTTCCA	CTAGGAGTTG	5760
	GAGATAGTGT	AATAAATGCC	AATGACACTC	TGAAGTTACA	GCGGATGGGT	GTTTTGCCAA	5820
35	AATTCATGG	TGGTACGAAA	AAGAAAAAAT	GGATGGAACA	AGTTACTGAA	AATCTTGGTA	5880
	AAAAAGCAGG	GGACTTCGGT	TCTAAAGCTA	AAAACACAGC	TCATAATATC	AAAAAAGGTG	5940
	CAGAAGAAAT	GGTTGAAGCG	GCAGGCGATA	AAATCAAAGA	TGGTGCATCT	TGGTTAGGCG	6000
40	ATAAAATCGG	CGATGTGTGG	GATTATGTAC	AACATCCAGG	GAAACTAGTA	AATAAAGTAA	6060
	TGTCAGGTTT	AAATATTAAT	TTTGAGGCG	GACTAACGCT	ACAGTAAAAA	TTGCTAAAGG	6120
45	CGCGTACTCA	TTGCTCAAAA	AGAAATTAGT	AGACAAAGTA	AAATCGTGGT	TTGAAGATTT	6180
	TGGTGGCGGA	GGCGATGGAA	GCTATCTATT	TGACCATCCA	ATTTGGCAAA	GGTTTGGGAG	6240
	TTACACAGGT	GGACTTAACT	TTAATGGCGG	TCGTCACTAT	GGTATCGACT	TTGGTATGCC	6300
50	TACAGGAACG	AACATTTATG	CTGTTAAAGG	CGGTATAGCT	GATAAAGTAT	GGACTGATTA	6360
	CGGTGGCGGT	AATTCTATAC	AAATTAAGAC	CGGTGCTAAC	GAATGGAAC	GGTATATGCA	6420
55							

	ATCAGGTGCT ACAGGTAATT TCGTTAGAGG AGCACACTTA CATTTCCAAT TGATGCAAGG	6540
	GTCGCATCCA GGGAAATGATA CAGCTAAAGA TCCAGAAAAA TGGTTGAAGT CACTTAAAGG	6600
5	TAGTGGCGTT CGAAGTGGTT CAGGTGTTAA TAAGGCTGCA TCTGCTTGGG CAGGCGATAT	6660
	ACGTCGTGCA GCAAAACGAA TGGGTGTTAA TGTTACTTCG GCTGACGTAG GAAATATCAT	6720
	TAGCTTGATT CAACACGAAT CAGGAGGAAA TGCAGGTATA ACTCAATCTA GTTCGCTTAG	6780
10	AGACATCAAC GTTTTACAGG GCAATCCAGC AAAAGGATTG CTTCAATATA TCCCACAAAC	6840
	ATTTAGACAT TATGCTGTTA GAGGTCACAA CAATATATAT AGTGGTTACG ATCAGTTATT	6900
15	AGCGTTCCTT AACAAACAGAT ATTGGCGCTC ACAGTTTAAC CCAAGAGGTG GTTGGTCTCC	6960
	AAGTGGTCCA AGAAGATATG CGAATGGTGG TTTGATTACA AAGCATCAAC TTGCTGAAGT	7020
	GGGTGAAGGA GATAAACAGG AGATGGTTAT CCCTTTAACT AGACGTAAAC GAGCAATTCA	7080
20	ATTAAC TGAA CAGGTTATGC GCATCATCGG TATGGATGGC AAGCCAAATA ACATCACTGT	7140
	AAATAATGAT ACTTCTACAG TTGAAAAATT GTTGAAACAA ATTGTTATGT TAAGTGATAA	7200
	AGGAAATAAA TTAACAGATG CGTTGATTCA AACTGTTTCT TCTCAGGATA ATAACCTAGG	7260
25	TTCTAATGAT GCAATTAGAG GTTTAGAAAA AATATTGTCA AAACAAAGTG GGCATAGAGC	7320
	AAATGCAAAT AATTATATGG GAGGTTTGAC TAATTAATGC AATCTTTTGT AAAAATCATA	7380
	GATGGTTACA AGGAAGAAGT AATAACAGAT TTTAATCAGC TTATATTTTT AGATGCAAGG	7440
30	GCTGAAAGTC CAAACACCAA TGATAACAGT GTAAC TATTA ACGGAGTAGA TGGTATTTTA	7500
	CCGGGCGCAA TTAGTTTTGC GCCTTTTTCA TTAGTATTAA GGTTTGGCTA TGATGGTATA	7560
35	GATGTTATAG ATTTAAATTT ATTTGAGCAT TGGTTTAGAT CTGTGTTTAA TCGCAGACAT	7620
	CCTTATTATG TTATTACTTC TCAAATGCCT GGTGTTAAAT ATGCAGTGAA TACAGCTAAT	7680
	GTTACATCTA ATTTAAAAGA TGGTTCTTCA ACTGAAATTG AAGTAAGTTT AAATGTTTAT	7740
40	AAAGGGTATT CTGAATCAGT TAATTGGACC GATAGCGAGT TCTTATTCGA CTCTAATTGG	7800
	ATGTTTGAAA ATGGAATTCC TCTTGATTTC ACACCTAAAT ATACTCATAC ATCAAATCAA	7860
	TTTACTATTT GGAACGGTTC TACTGATACG ATAAATCCAC GATTCAAGCA CGATTTGAAA	7920
45	ATATTAATTA ATTTAAATGC GAGTGGAGGA TTTGAACTGG TTAATTATAC AACAGGTGAT	7980
	ATTTTTAAGT ACAACAAAAG TATAGATAAA AACACTGATT TTGTTTTAGA TGGTGTGTAT	8040
	GCATATCGAG ATATAAACAG AGTGGGAATT GATACAAATA GAGGCATTAT AACATTAGCG	8100
50	CCAGGTAAAA ATGAATTTAA GATTaAAGGA GACGTCAGTG ATATTAAAAC TACATTTAAG	8160
	TTTCCTTTTA TTTATAGGTA GGTGATTTAA TGGATTATCA TGATCATTTA TCAGTAATGG	8220
55		

ATTATGAACT GAATGAAGCT AGGTACATCA CCTTTACAGT TTATAGAACT ACTCATAATA 8340  
 GTTTTGTGTTT TGATTTATTG ATTTGTGAAA ACTTCATAAT TTATCATGGT GAAAAATACA 8400  
 5 CAATTAAGCA GACAGCGCCA AAGGTTGAAG GTGATAAAGT TTTTATTGAA GTTACGGCAT 8460  
 ATCACATAAT GTATGAATTT CAAAATCACT CAGTGAATC AAATAAGCTT GATGACGACA 8520  
 GTAGCGAAAC TGGTAAACG CCAGAATACT CTTTAGATGA GTACTTAAGA TATGGATTG 8580  
 10 CAAATCAAAA AACGTCAGTC AAGATGACCT ATAAAATAAT TGGAGATTTT AAAAGAAAAA 8640  
 TACCAATTGA TGAATTAGGT AATAAAAAATG GCTTAGAATA TTGTAAAGAA GCAGTAGATT 8700  
 TGTTTGGTTG TATTATTTAT CCAAATGATA CGGAGATATG TTTTATTCTT CCTGAAACAT 8760  
 15 TCTATCAAAG AAGCGAAAAA GTAATAAGGT ATCAATATAA TACTGATACT GTGTCTGCTA 8820  
 CTGTCAGTAC GTTGAATTA AGAACAGCTA TAAAAGTTTT TGGGAAAAAG TACACAGCCG 8880  
 20 AGGAAAAGAA AAATTATAAT CCTATTAGAA CAACTGACAT TAAATATTCA AATGGTTTTA 8940  
 TAAAAGAAGG TACTTATCGT ACCGCAACAA TTGGGTCTAA AGCTACTATT AACTTTGATT 9000  
 GCAAGTATGG TAATGAAACA GTTAGATTTA CAATAAAaAA GGGCTCTCaA GGTGGAATAT 9060  
 25 ATAAGTTGAT TTTAGACGGC AAGCaAATTA AGCaAATTTT TTGTTTTGCT AAGTCGGTTC 9120  
 AGTCTGAmAC AATAGATTTA ATaaAAAAATA TTGATAAAGG CAAGCACGTT TTAGAAATGA 9180  
 TATTTTTTrGG AGArGrCCCC AAAAATAGAA TTGATATATC TTCAAATAAA AAAGCTAAGC 9240  
 30 CTTGTATGTA TGTTGGAAC TAAAAATCAA CAGTCTTAAA TTTAATTGCT GACAACTCAG 9300  
 GTCGCAATCA ATACAAAGCA ATTGTTGaCT ACGTCGCAGA TAGTGCAAAG CAGTTTGGA 9360  
 TTCGATATGC TAATACGCAA ACAAATGAAG ATATCGAAAC ACAGGATAAG CTGTTAGAAT 9420  
 35 TTGCAAAAAA GCAATAAAT GATACTCCTA AGACTGAATT AGATGTTAAT TATATAGGTT 9480  
 ATGAAAAAAT AGAGCCAAGA GATAGCGTAT TCTTGTTC TGAATTAATG GGATATAACA 9540  
 40 CTGAATTAAA GGTTGTAAA CTTGATAGGT CACATCCATT TGTAACGCA ATAGATGAAG 9600  
 TGTCTTTCAG CAATGAAATA AAGGATATGG TACAAATTCA ACAAGCGCTT AACAGACGAG 9660  
 TTATTGCACA AGATAATAGA TATAACTATC AAGCAAATCG TATAAATCAT TTATACACTA 9720  
 45 GTACTTTGAA TTCTCCTTC GAGACAATGG ATATAGGGAG TGTATTAATA TAATGGCAAC 9780  
 AGAAGAAGTT AAAATCAAAG CGCTACTTGA AAACGATAAA C 9821

(2) INFORMATION FOR SEQ ID NO: 471:

- 50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1017 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 471:

5 TTATTGTTTT CCAAATGGTC TTCTCGGTTA TTAAACCATC GGTTCATC CTACGTGCTT 60  
 CACTTTCTAT CtATTAATTC AaTTTCAGCT TGACCACCCG CTGTATAAAG GGTCAAAGTT 120  
 GCTAATCGAT AGCGTCTCAT TATAGGACCA ACATCGATAT CAATATTTTG AATACGAAAA 180  
 10 TATGGTATTA CCTTTTCATC CAAAAATAAA ATGCCGTTTC GTACACGCAA ATGGTGTTTT 240  
 TCAAATGCAT ATCTGCAGTG CTTATATCGA TAGACTGGCG CTATAACAAG CGTGAAAACA 300  
 GCAACAAGTA ATATTATAAT CACACTACTA ACAATGGATA AATGGTTATC TAAAATCTCC 360  
 15 CAAAATAGCC AGTTCAAAAT ATTAAATGCG ATTAAAAGTA CAAGCGCTAT GGGTATCCAA 420  
 AACAGCACAG CACTTAACCT CATCACTTTT TTAGCGTGTG GTGACATAAA ATTATAATCC 480  
 CTCATCATTT TCACCTCTTA AATACCATGA TTTCAATTTA TTTGCATCTT CACTTCTTGC 540  
 20 GTATTTTAAAG TTAATCGATT GGGCTCCCAC ACCTTTAGCA ACAATAAAGC TAAAATTATT 600  
 TAAATTGTTG CGTTTAAGTA ATGTATGTTG CCAAGTGTC AATCCTATAA TGTGATGCGC 660  
 TTTAAAATAA TAAATATTTT GTTTCAATAG CTCGAAATTC TGGATAGTAA TTTGTTCTTC 720  
 25 TGTCAITTTA AAACCCGCAT GTTTGACATA AAGATATCCT TTGATCACAA ATAAACCAAT 780  
 AATGACTATT GTTATAATCG TAAATAACAA TAATAATTGA TTCCAAAAAT AACAGCCTAT 840  
 30 ACCTGCCATA GCTATGACAA TAATACTAGG TATTAAAAAG TGTCTGTGGA AACCTGACAA 900  
 AGGCATACCT TCATTAACTT GTTGATAAGA TAAATCTGGT ACTAAATTCT GGATAATTG 960  
 ATATGCTTTG TCTCGTTTAA TAAACGGCaA TATCGGCACA CTACCTGAAT CATTGTC 1017

## (2) INFORMATION FOR SEQ ID NO: 472:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6806 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 472:

45 TAGGATTGAT TAATCCTTTC TTCAAATGA TGAATGTGTA ATGTTAAATA TATAATTTCA 60  
 GACTCACTGA TATTAACATC AAATTGTTTT TGTATCATAG TTAATAATTTT ATATGCTGTG 120  
 50 TTATAGCAAA TCGGATAGTG ATTTTAAATC ATAGACACAA AATCATCTTG TGCATGTATA 180  
 TATTCTTTTC TTCTTAATCG GCGAATTAAA AATTGTACGT GCCTTATAAA ACGTTGGTAT 240



	TTAATAACAT TATTnGATCA AGGTCATCTC ATGCATAGAT AAATCTTCTG TATTaGATGC	360
	AATATGTAAT GCAATAAATC CTATCTCATC TTCAGGaAAA TGaCATCTA ATGCTGCATT	420
5	TAACTGATCA ATCACCTGTT TAGCAATATG ATATGCATCA CTATATAACT GCATAGTTTC	480
	CATAACAAAT GGATTGCTAA TAACTTGATT TTGTTTTAAG CGTTTATAAG CAAATATAAT	540
	ATGATCCGTT AATGAAACTA CAAGTTGTTT TGAATCAACA TTCATCGCAG TATTAGAAAT	600
10	AAAATTCAAC GAATCAATAA TTACTTGTA TACATTATCA TCAGCGATT CAACTAACT	660
	TTTATAATGT GCTTTTGTG GCTCACTTTC TAATTTATAA ATTTTCTCTA TTGTAATAGT	720
	TTGGTCGTTT AACGCCATT CCTCTTTTTT GTTAAACCA ATACCTTTAC CGATTAAAT	780
15	AACTTCTTGA TCATTATTAG TACATACTAC GACATTGTTG TTCAATGTT TAGTAACAA	840
	ATATTCTCCC ATTATCATCA CCTATTTTTT TATTATTAAG ATTATATATC GGAAATGTCT	900
20	AACTTGAAAG TATAATAATT TAAATACTTG TGTCTATTA CACGCCTATC CTATATGATA	960
	TATCTTAATT TAATTTGAT GTCTCTCAA GTGGAATAAC TATAATAAAA AAATCTGACT	1020
	CCCATATTTA CAAATAATTC TATTTATTTA TATATTATCT GAATTAATAC TCAATACAAA	1080
25	CTAAAACGTA CTATTAAATT GTGCAAAGCT AAAACAAATT TATATTCATC TATCCAACAA	1140
	TATGTCTTAT CAATGGTATA GTCTTGCAC ACCAATGGAG GAAATAATC TCAACCTTAC	1200
	TATATTAATA TATAATCAAA TCTTAGATTA ACTAGTGTA TGATACAGAT GATAATTGAG	1260
30	TACAAATTTA AAACCCTGAG ATTTTCGCTT TAATTTGAAA ACCTCAGGGT TTATTTGATT	1320
	TTTATATAAT GAATCGTTAC ATTAAAAATA TTTATTTATC AGAGTTCTTA TATTTGTTAG	1380
	CGCCCCAAGC ACTAATTCCA AATAAGTTAA TTTCTAAGTT TTCAGGTTTA AAGACAGGGT	1440
35	TCTTGCCTTC TTTTTTCTGC TTTTGATAAT CTTTCATCAA TGCAAAAGCT ACATTGGACA	1500
	GTCCTATAAT GGAAATAATG TTTACAATTG CCATTAAGCC CATAAATAAG TCTGCCGTAT	1560
40	TCCATACTGT TTCTGTTTTT ACAACTGCAC CGACAAAGAC AAGTACTACA ACAAGACATC	1620
	TAAAGATAAA TAATATTACA CGGTTTGTG ATAAAAATTC AATATTAGAT TGACCGTAAT	1680
	AGTAATTACC TACAACAGAT GAAAATGCAA ACAGTGTAAC tGCTaTTGTT AAGAAAATAC	1740
45	CTCCAGCAGA ACCTAAATGC TCATTAAGTG CTGATTGAGT AACTGCAACA CCTTGAGGTG	1800
	CGTTATCACC AAATTTCACT CCTGAATATA GTAAATCAT GATTGCAGTT GCTGTACAAA	1860
	CCAACATTGT ATCAAAGAAC ACACCTAATG ATTGGATTAA ACCTTGCTTA ACAGGGTGTG	1920
50	GTACGGCAGC AGTTGCCGCT GCATTGGCG CAGAACCCAT ACCAGCTTCG TTAGAGAATA	1980
	AACCACGTTT GATACCTTGA AGAACCGCAG CACCTACAGC GCCACCAGTT ACTTGTTCTGA	2040
55		

	GCAATATTAC TAAAACCATA CCAATGTAAA TGATAGCCAT AATCGGTACA ATTAACGAAG	2160
	ATAACGTAGC AATACTACGT ACACCACCAA ATATAATAAT AGCTGTTACG ATTGCTAAAA	2220
5	TAATACCTGT GATTACTGGA CTAATATTAT ATTGCGTATT TAACGACTCC GCAATTGTAT	2280
	TAGATTGCAC TGTGTTAAAT ACAAATGCAA ATGTAATTGT AATTAAAATC GCAAATACGA	2340
	TACCTAGCCA TTTTGTATTT AAACCTTTAG TAATATAGTA AGCTGGACCA CCACGGAATC	2400
10	CACCATCTTT ATCATGTACT TTATAAACCT GAGCCAAAGT CGCTTCTATA AATGCACTCG	2460
	CTGCACCTAT AAATGCAATA ACCCACATCC AAAATACTGC ACCTGGACCG CCTAAAAACA	2520
	TCGCAGTCGC AACACCAGCA ATATTACCAG TACCAACTCT CGAACCAGCA CTAATCGCAA	2580
15	ATGCTTGGAA TGGCGAAATA CCCTTCTTAC CATCTTCTAA AGTTTCTGGA CGTTCTACTA	2640
	AAGCTCTAAA CATTTTCAGGT AACATTGCGT ATTGAACGAA TTTAGAACTA ATCGTAAAGA	2700
	AGAATCCAGC TGTCAATAAT AGACCAATTA AATATTGAGA CCATATTAAA TCGGTACCAA	2760
20	CATGGACAAA TTCTTTAAAC CATCCAGGTA TTAAACTATC GAAATCTTTC AAAATAAACC	2820
	CCTCGCATCC TCTACATGAA TCATGTACCT TCTATAAAAT TAGACCGAAT TGAACCTTCA	2880
25	GTAAATATAG AGATACATCA TCATTCTTA TACAATACAA GAGATTTATA TTAGTTTGGT	2940
	CAAAGTATAT CGCTAATTTA ACGATAAGTA CTTGGTCAGC ATTTAATATA AATCCCTTGA	3000
	ATTTAGTCAA AATTTAACAT TACTGTATTT TATCATTTAA TTTCGTGATT GCATATAGTT	3060
30	TTTAGCTAAT ATACATGTCT ATTACTTCAC CAAAATCATC TGTATCTACA ATGAATGAGC	3120
	CATTTGTATA TTGTTTCAGAT TTATGAATAT CATTAAATTAA ACCATGTTCT TCATTTGATT	3180
	TTGAATATAA TGTATATTGA CTATGTTTAC CTGTCACTAC ATGTGCAGCT ACAATACGAT	3240
35	GTGGATTTTT CTTTAATTCT TTTAATAAAG TTATTCCaCG TTGTGCTCTT TTAGCAACTT	3300
	GTAAGATTTT AAAACTAATA CGTTTTAACG AGCCGCGTTG TGTGGCCATC AATATAGTAT	3360
	CATTTTCAGA AACACCTTCT GTCATAACAA CGAAATCTTC AGCTTTAAGA TTTATTGATT	3420
40	TAACACCAGC TGCCCTTAAT CCGGTATCTG ATAGTTCACT TGTATTATAC GTTAATGACA	3480
	TACCTTTATT AGTAATGACG GTAATTAAAT GATCTTTTTT AAAGCGCATA AACTAATCA	3540
45	AATCATCATT TTCTTTAACT TTAGTAGCAA TTAAAGGTTT ATTAAAACGC GTTGTTTTTAA	3600
	ATAGAGGCAC TGTACTTTTC TTAATCATGC CATTTTGAGT CGCAAAAACA TAAAATGCAT	3660
	CTGTATTAAA GTCCTTTTCA TTAAAGACAT TAATAACCAC TTCATCTTCT TCGATAGGAA	3720
50	CTATTTGTGA TACATGTTGT CCCAATTCTT TCCAACGAAT ATCTGCTAAT TTATGAACCG	3780
	GTATAAATAG ATAACGACCT TTATTTGTAA ATACTAGTAC GGTATCTTGC GTATTTACTT	3840

55

	TAAAGCTACG AATAGAAGTA CGTTTAATAT ATCCATGACG TGTCATACTT AAAATAACTY	3960
	CTTCACTAGG CACCATAACT TCTTTGTCAA TTTTAATTTT TTCAATTTCT GCTTCAATTA	4020
5	AAGACAGTCG TTCAGATTTG AATTTCTTTT TAATTTTCATT CAATTCCTTCT TTTATGACAT	4080
	TCAATAATGC ATCATGGTTA TCAAGAATAT GACGTAATTG TTTGATTAAT GCTTCAAGTT	4140
	CTTTATGTTT ACCTTCAAGC GCAACTATGT CAGTATTTGT TAAACGATAT AACTGTAACA	4200
10	TTACAATTGC TTCAGCCTGT TCTTCTGTGA ACTCGTATAC TTCGATAAGG TTTTCTTTAG	4260
	CGTCACGCTT GTTTTTAGAG CWACGAATCA ATTCGATTAC TTTATCTAAA ATTGACAACG	4320
	CTTTAATCAA ACCTTCAACG ATATGCATAC GTTTTTCTGC ATTATCTAAT TCAAACCTCG	4380
15	TTCTATTTGC AACAACTCA ATTTGGTGAT TCAAATAACT ATCTATAATT TGACGAATAC	4440
	CCATCAATTT TGGACGACCA TCACTAATAG CGACCATGTT GAAATTATAT GAAATCTGTA	4500
20	AATCAGAGTT TTTATAAAGA TAATTTTGA TTGATTCAT GTTCACATCT TTTTCAATT	4560
	CAATTGCTAT TCGTAAACCA GTTCTATCAG TTTTCATCAG TACTTCAACG ATACCATCGA	4620
	CTTTTTGTC AGCACGTAAT TCATCGATAC GTTTTACTAA GCTACTTTTG TTTCACTTCAT	4680
25	ATGGAATTTT AGTAATAATT AACTGTTTAC GTCCATTGCG TAAAGTTTCT TCTTCAACTT	4740
	TAGAACGAAC TATAATTCTA CCTTTACCTG ATTCATAAGC TTTTTTAATA CCATCAATAC	4800
	CTTGAATAAT ACCACCAGTT GGAAAATCAG GACCTTTAAT ATATTTTCATT AATTGATTGA	4860
30	CTGTAATATC CGGATTATCA ATATATTTAA GTGTTGCTTG AATCACTTCA GCTAAATTAT	4920
	GTGGTGGTAT ATCTGTCGCG TAACCTGCAG ATATACCTGT AGAACCATTG ACTAGTAAGT	4980
	TAGGAAATCT TGATGGCAAT ACCATTGGTT CGAGTGTCTG ATCATCATAG TTTGGAATGA	5040
35	AAGAACTGT CTCTTTATTA ATATCACGTA ATAACCTTCT AGCTAGTAAG CTTAACTTAG	5100
	CTTCAGTGTA ACGCATTGCC GCTGGCGGAT CATTATCGAT ACTACCATTA TTACCATGCA	5160
	TTTCTATTAA GACATGTCGT AACTTCCAGT CTTGACTTAA ACGGACCATT GCTTCGTACA	5220
40	CTGAGGAGTC TCCATGTGGA TGATATTGAC CAATAACATC ACCGACTGTT TTCGCACTTT	5280
	TACGGAAATT TTTATCGTGT GTATTACCAC TTGAATACAT TGCATATAAA ATACGACGTT	5340
	GTACTGGTTT TAAACCATCA CGAACATCTG GCAATGCACG CTCTTGAATA ATATATTTAC	5400
45	TATATCTTCC AAAGCGATCA CCTAAAACAT CTTCAAGTGA TAAATCTTGA ATTATTTAC	5460
	TCACTAGATT TCCTCCTCAT CAAATTGATC ATTTTCAAGC ACTTGACTT CAGAAATTATC	5520
50	TAAAATACTT TGGTCCTCTT GCATACCAA CTCAACATGC TTTTCAATCC ATTCACGTCT	5580
	AGGTTGTA CT TTGTCACCCA TTAATGTTGT TACACGTTTA GATGAACGCA CTTCATCTTC	5640

55

AGGGTTCATT TCACCCAAAC CTTTGTAACG TTGTAACGTG AaGCCTTTAC CAAGTTCTTT 5760  
 TTGCAATTTA TTAAGCTCTT CGTCTGTCCA AGCGTATTCA ACTCGCTTTG TTTTGCCTTT 5820  
 5 ACCTTTTTCC AATTTATAAA GTGGAGGTAA AGCAATAAAT ACACGACCTG CTTGAACAAG 5880  
 CGGTTTCATA TATTTGAAGA AGAATGTAA CAATAGCACT TGAATATGCG CACCATCAGT 5940  
 ATCAGCATCA GTCATAATAA TTACACGATT ATAATTACTA TCTTCAATTT TAAAGTCAGT 6000  
 10 ACCAACGCCT GCCCCGATTG TGTGGATAAT TGTATTAAAT TCTTCATTTT TAAAAATATC 6060  
 TTCTAGACGT GCTTCTCTG TATTAATTAC CTTACCACGT AATGGTAATA TCGCTTGGA 6120  
 TTTGCGGTCT CGTcCAAGTT TTGCTGAACC TCCCGCAGAA TCACCTTCGA CTAAATACAA 6180  
 15 TTCAATTTTT TCAGTGTTTT TACTTTGTGC AGGTGTTAAT TTACCAGATA GCAAAGTGTC 6240  
 TTTACGCTTG TTTTTCTTAC CTGAACGAGC ATCTTCACGA GCTTTACGTG CAGCTTCCCT 6300  
 TGCTTGTTGT GCTTTAATCG CTTTTTTCAC AAGTGATTTA GACAATTGTC CTTTTTCTTC 6360  
 20 TAAATAGAAT GGCAATTTGT CTGCAACAAC TGAATCAACA GCACCTCTAG CTTCAGAAGT 6420  
 ACCCAATTTA GATTTGTTTT GTCCTTCAAA TTGCAATAAT TCTTCTGGAA TACGAACAGA 6480  
 25 CACAACAGCT GTTAAACCTT CACGAATATC ATTACCATCT AAGTTTTTAT CTTTGTTTTT 6540  
 AAGTTCATTA ATACGACGTG CATAATCATT AAATACACGT GTCATTGCTG TTTTAAAACC 6600  
 AACTTCATGT GTACCACCAT CTTTAGTACG TACATTATTT ACAAACCTTA AAATACTTTC 6660  
 30 TGAATATTGA TCATTATATT GGAAAGCTAC GTCTACCTCT ATACCATTG CTTACCTGA 6720  
 AAATGTAGCC ACGTCATGCA AAACCTCTTT TCCTTCATTG ACATAACTAA CAAACTCTTT 6780  
 GATTCCTTCT TATAATGGTA TGTCTT 6806

35 (2) INFORMATION FOR SEQ ID NO: 473:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1716 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 473:

GGGGCAnAAA TTCCAATACA CTCATTACCA AATATATACA CCACTTCCTG CTACAAGTnn 60  
 TTTTACTTGA TCTTGGTCTT TTCCGCAGAA AGAGCAITTC CAAATTTTCT TCATCTTCCA 120  
 50 TTGAATTTAA ACATTCTTTT TACACCCCTA TTCGTAAAG ACTATACTAG ATTGGATGTT 180  
 ACAATGCAAC ATATTAAACAT ACAAACCTTT TGCTTAAAGA ATAGTAGCAG ATACATAAGC 240

55

EP 0 786 519 A2

CCTTCAACGA ACTTTGCGTT ATCTCTTAAT AAATCGATAA CTTTTTGGAT ACGAACATCA 360  
 TTTTAAATGA TATCAGTATT ACCTAAAGTA TTTTGGATAT CTTCAACTGA GATATTAAAT 420  
 5 TGTTTACTCA TTTTTTCTAA TTCTTTATCG ATATCTTCAT CAGTAGCTTC GATTTTTTCA 480  
 GCTTCAGCGA TCGCAGTTAA AGTTAAGTTA GTTTTAACAC GTTGTTCTGC ATCGTCTTTC 540  
 ATTTGCTCTC TTAATTGAGT TTCATCTTGA CCTGAGATTT GGAAGTACGT TTGTAAATCT 600  
 10 AAACCTTGTT GTTGAATTCT TTGTGCAAAT TCAGACACCA TACGATCTAA TTCAGTATTA 660  
 ACCATTGCTT CAGGAATATC GATTGTGTGA TTATCAGTAG CTTTTGTAAT CGCTTCTTCT 720  
 TTTTCAACAT TTTCAGCATC TGTAGCTTTT TGTTTACGTA AACGTTTACG TAAGTTTTCT 780  
 15 TTGTACTCGT CTACTGTATT TGCTTCTGCA TCTAATTCAT TAGCAATTC ATCTGTTAAT 840  
 TCTGGGACTT CTTTAAATTT AATTCGTTA ACTTTTGTTC TGAAAGTTGC TTCTTTACCG 900  
 GCTAATTCCT CAGCATGGTA TTCTTCTGGG AATGTTACGA CAACATCTTT TTCTTCGTCA 960  
 20 ACTTTCATAC CTTCTAATTG CTCTTCGAAA CCAGGTATGA ATGAACCTGA ACCGATTTCT 1020  
 AAATCGTAAC CTTTACGCTG TCCACCTTCG AATCTTCTC CGTCAACTGA ACCACTAAAG 1080  
 25 TCGATGTTAA CTGTGTCGCC ATTTTCAACA ACACCATCTT CTTTAAACGAC CATTTTACGCT 1140  
 AAATGTCCTA AGCTGTGGTC AATCGCTTCT TGTAATCAT CATCAGATAA TTCAGTTTCT 1200  
 TGTTTTTCAA TTTCAAGACC TTTATAGTCT CTTAATTTAA CTTCTGGCTC AACTGTAAT 1260  
 30 GTTGCTTCAA AAATGAAATC TTTACCTTTT TCAATTTGAG TAACACTTAC TTCTGGTTGT 1320  
 GCAACTGGTT TAATATCAGT TTCGTCAATT GCTTCACCAT AAGCATCTGG TAATAAAATG 1380  
 TCGATAGCAT CTTGATATAA TGCTTCTACA CCAAAGCGTT GTTCAAAAAT TGGACGTGGC 1440  
 35 ACTTTACCTT TACGGAATCC AGGTACGTTA ATTTGTTTAA CCACTTTTTT GAATGCTTGA 1500  
 TCTAACGCTT TGTTTACTTT TTCTGCAGGA ACAGTAACAG TTAATAAACC TTCGTTACCT 1560  
 TCCTTTTTTT CCCAAGTTGC TGTCATGTAT ATATACCTCC ATGATTAAT AATTTATTTT 1620  
 40 TTCAACTTCC CTATTATATC ATACGTCTAT TCCCTATACA AACATTGAAA TCACAACGTT 1680  
 TATATATTTG TAAATCAACT TTTTTCGTCA AAATA 1716

45 (2) INFORMATION FOR SEQ ID NO: 474:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 795 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

55

	TGACCAAGTA CGTTTCGAAG TTGCCATTAA AGCATTAAAC CCATCATTGA AAGCATTTCGC	60
	ACCTGTACGT GAGTGGGCAT GGAGTCGTGA AGAAGAAATC GATTATGCAA TTAAACATAA	120
5	TATCCCTGTA TCAATCAACC ATGATTCACC TTATTCTATC GATCAAAATC TATGGGGCAG	180
	AGCGAATGAA TGTGGTATTT TAGAAGATCC TTATGCTGCG CCACCAGAGG ATGCGTTCGA	240
	TCTAACAAAT GCTTTAGAAG AAACACCAGA TACTGCTGAT GAAATCATT TAAACGTTTGA	300
10	TAAAGGCATC CCAGTTCAAA TTGATGGCAA AACATATGAA TTAGACGATT TAATTTTAAAC	360
	GTTGAATGCA TTAGCTGGTA AGCATGGTAT CGGAAGAATT GACCATGTAG AAAATAGACT	420
	TGTAGGTATC AAATCAAGAG AAATTTATGA GGCACCTGCT GCAGAAGTTA TTTTAAAAGC	480
15	GCATAAAGCA TTAGAAACGA TTACGTTAAC GAAAGATGTC GCACACTTTA AACCAATCAT	540
	TGAGAAGCAA TTTGCTGAAC AACTATACAA TGGACTTTGG TTCTCACCTT TAACTGATAG	600
	CTTGAAATTA TTTATTGATA GTACTCAGCA ATACGTAAGT GGTGATGTCA GAATTAAATT	660
20	ATTCAAAGGT AATGCCATCG TGAATGGTAG AAAATCACCT TACACATTAT ATGATGAAAA	720
	ATTAGCAACT TATACAAAAG AAGATGCATT TATCAAGAC GCTGCTGTTG GCTTTATCGA	780
25	TATCTATGGT TTACC	795

(2) INFORMATION FOR SEQ ID NO: 475:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 887 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 475:

	CGATTGAAAG AAGACGGTTC AGTTGAAAAG TTTCCAAAGC CAGTAATTAG CCAACAACCA	60
	GAAGGATATA CGAGTCATTT TAGAGATCCT AAAGTTTTTA AATATGATG GaAATATTAT	120
40	GCAATCaTTG GTGmCaAAA TAATGATCaG CaAGGTCGAT TATTACTTTA TAATACTGAA	180
	GATATAATTA ATTGGCATT TTTAGGTGAA ATAAATACAG AGTTGGATGA TTTTGGATAT	240
45	ATGTGGGAAT GCCCAGATTA CTTTAATGTA GATAATCAAG ATGTCATACT TATTGTCCA	300
	CAAGGTATTG AACCAAAAGG CGATCAGTTC AAAAATATTT ATCAAAGTGG TTATATACTT	360
	GGAAAGTTTG ATATTGAAAA GTTAACATAT GAACATGAAA ATTTTGTCTGA GCTTGATAAT	420
50	GGTTTTGATT TCTATGCACC TCAAACATTT TTAGATGAAA AAGGCCGACG AGTACTAATT	480
	GGATGGATGG GGTTACCGGA AATCGAATAT CCTACTGATA ATGAAGGATG GGCCCATTCG	540

GCGTTGAAA AATTACGTCA CAATAAAGAG ACAGCATTtA GGctACGCAA ATAAATTTAC 660  
 TCGAAAATTA CATCCGTATG AAGGTAAACA GTATGAATTA ATCATAGATA TTTTGGATAA 720  
 5 TGATGCTACC GAAGTGACTT TTGAATTACG TACATCTAAG ACTTCTTCAA CATTAAATTGC 780  
 TTATAACAAG CGTGAATAA AAATAACATT AGATCGCAGC GACAGTGGTT TATTGCCGAC 840  
 AAATGTTGAA gGTACGACGC GTAGTACGAT ATTAGACACG CCATTAA 887

10 (2) INFORMATION FOR SEQ ID NO: 476:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1183 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 476:

20 TTGGAAnCAA AAAACCATTG GTAAACCGTG TtnAACCGGA TTTCCGATGG ACCTTTTAAA 60  
 ACnACCAAAT AGAAAGCTTT GATAAAAGGT AATTATGGTA CTGATAACAA ACAAGTTCAA 120  
 25 AAACATCATG ATTTAGTACG TATGCTTTTG ATGGATCAAG ATGGTTTTTT AACTGAAAAT 180  
 AATAAAGTTG ATCATTTTCAT TGATGGAAAT GATTTATATG ATCAAGTTTT AAAAGATATT 240  
 AAAAATGCAA AAGATATATC CATTTAGAGT ACwATACTTT CGCTTwAGAT GGTTwAGGTA 300  
 30 AAAGAATTTT ACATGCTTTA GAAGAAAAAT TGAAACAAGG TCTAGAAGTA AAAATATTAT 360  
 ATGATGATGT TGGATCTAAA AATGTTAAGA TGGCAAATTT TGATCATTTT AAATCGTTAG 420  
 GTGGAGAAGT TGAAGCATTT TTTGCTTCAA AATTACCGTT ATTGAATTTT AGAATGAATA 480  
 35 ATAGAAATCA TAGAAAAATC ATCGTAATCG ATGGTCAACT AGGTTATGTC GGAGGATTTA 540  
 ACATTGGTGA TGAATATCTm GGATTAGGAA AATTAGGATA TTGGAGAGAT ACGCATTTAC 600  
 GTATACAAGG GGATGCGGTT GATGCACTGC AGTTGCGATT TATTTTAGAC TGGAATTCGC 660  
 40 AAGCGCACCG TCCACAATTT GAATATGATG TTAAGTATTT CCCTAAAAAG AACGGACCAT 720  
 TGGGCAATTC ACCAATTCAA ATAGCTGCAA GTGGCCCGGC TAGTGAAGTG CATCAAATTG 780  
 AATACGGTTA TACAAAAATG ATTATGAGTG CAAAGAAATC TGTATATTTA CAATCACCAT 840  
 45 ATTTCAATTCC GGATAATTCA TATATaAATG CCATTAAAAT TGCTGCTAAA TCAGGTGTAG 900  
 ATGTACATTT AATGATTCCA TGTAAGCCAG ATCATCCATT AGTATATTGG GCGACATTTT 960  
 50 CAAATGCCTC TGACTTATTA TCAAGTGGTG TTAAAATTTA TACGTATGAA AATGGATTTA 1020  
 TACATTCTAA AATGTGCTTA ATTGATGATG AAATCGTATC AGTGGGCACA GCAAATATGG 1080

55

CTAAAGATTT AAGGGTGGCT TATGAACATG ATATTACAAA ATC

1183

(2) INFORMATION FOR SEQ ID NO: 477:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2332 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 477:

GGTTATATGT TTGGTATTAC TCATTATATT GATTGGAGGT TGTGTCATTA TGACAAAAAC 60  
 15 AAATGGTCGA AACGCTCAAA TTAAAGAAAA TTTCAACAAA ACATTAAAGTG TATATCTAAC 120  
 CAAGAATCTC GATGATTTTT ACGATAAGGA AGGTTTTTCGA GATCAAGAAT TTGATAAAAG 180  
 20 AGATAAAGGG ACTTGGATTA TTTATTCTGA AATGGTTATC GAACCAAAAG GGAAATAATAT 240  
 GGAATCGAGA GGAATGGTGC TCTATATCAa TCGCAATACT AGAACmACGA AGGGTAATTT 300  
 TATTGTCACC GAAATAACTG AAGATAGTAA AGGATATTCA CGTAGTAAAG AAAAAAATA 360  
 25 TCCTGTCAAG ATGGAAAATA ATCGAATTAT TCCAACAAAG CCTATACCGG ATGACAAGTT 420  
 AAAAAAGAG ATTGAAAAC TTAAGTTCTT TGTACAATAT GGAAATTTTA AAGATTTTAA 480  
 AGATTATAAA AATGGTGATA TTTCATATAA TCCTAATGTG CCAAGTTATT CTGCAAAGTA 540  
 30 TCAATTGAAT AATGATGACT ATAATGTTCA ACAGTTAAGA AAACGATATC ATATTCCAAC 600  
 CAAACAAGCG CCCGAATTAA AATTGAAAGG ATCCGGCAAT TTAAAAGGCT CATCCGTAGG 660  
 ATCTAAGGAT CTAGAATTTA CGTTTGTAGA AAATCAAGAA GAGAATATCT ATTTTTCAGA 720  
 35 TTCGGTCGAA TTTACACCTA GCGAGGATGA TAAATCATGA GTCAAACGGA ATATCAAATA 780  
 AAATCTGGCA ATATAAAAGG TAACTCTGAA GAAACAAGTA CAGTATCTAA TATAAGTTAT 840  
 GAAATAGAAA ACGCAAATAA CAGTGGTTTA AAACAAAATA AAATTGATAA ACAAATTAAA 900  
 40 AAGTTACAAG AAAAAAATAA ATTCCCTAAA AATCTTTTCAT ATCTTAAAAG TTATACGGAC 960  
 CCCAAAACAG GCACGACTAC AAGCGCCTTT TTAAATAAAG ACACTGGCAA AGTTACTTTA 1020  
 GGTATGACAG GTACTAATGT ACACAAAGAC GCAATATTAA AACAAACATT TGGTGTTCCT 1080  
 45 TCTTATCAAG GATATATAGA CGTGAGTGAA ACgCTaAAAG ATATTGGGGC CGATGTCAAT 1140  
 ATTGGCCTTC ATTCCGTCAC AGATAAAGAT CCACATTATA AAAATACCCA AGACTTTATC 1200  
 50 AAAAATATCA AAAAAGACTA TGATATTGAT ATTATTACCG GACATTCGCT GGGCGGTAGA 1260  
 GATGCGATGA TTTTAGGTAT GAGTAATGAT ATTAAACATA TCGTTGTGTA TAATCCAGCT 1320

55



EP 0 786 519 A2

ATTGAAAAGT ACGATGGTCA CATTGTAAGA TTTGTGTCTG ATGAAGACGA ATTAGATGCA 1440  
 GGTGTCCGCA ATCATTATA TGAACTGCT GGAGAAAAA TAGTACTTAA AAATGGAGAA 1500  
 5 GGCCATGCAA TGAGTGGTAT TTTAATGAGC AGAACACAGG CTATAATCTT AGCTGAATTA 1560  
 AACAAAGTTA AAGGCTACCA AGACGAAAAT AATAAAGCAT TAAAATCCGT TCGTAAACAA 1620  
 ACGAGGCATA GATTACATAA AGTAGAGACG TTAAGAGCGA ATTGGATTCA AACACGGGT 1680  
 10 GGATCACTCT CTTCTCCCw ACAACAATTA TTAGAAGCTT TAACAGCACT AACCATTGCC 1740  
 GAAGGCTTAA ATCAATTAGT GAATGAAGAA AGCCAACATT TGAAAAAATG TATCACGCGA 1800  
 TGGCACATAA ATTTGGAGAC AACTGGAAAA AAGCGCAAGA AGTTGGAAAT GAAATTGGTG 1860  
 15 AAAAATTAAC CTCTGAAGAG GTTATAGATG rATTAAGAAA AGGTGGCGCG TATGAAAGTa 1920  
 AACTTGAAAC AGATCCCCAA AGAAAAATTG ATGATAAGAT AAAGAAATTA AATGATGTTT 1980  
 ATAAAAATG TAATGGCTAT ATCGCAAAAA TTAAACAGAG TATCGAAGCA ATTGTTTCTA 2040  
 20 ATGACCAAAT GTTAGCGAGC CAGATTGATG GGATGATGTA ATGTTTACTA CGTATAAnAA 2100  
 TATTAATGAA CTTGAAAATG CCTATGATGA AGAAAGAAAA CAATTGAATG ATGCATTCAA 2160  
 TCAAATTGAT GAATTAAGAC ATCAAACACG CAAGAmATGT GAACAAATGT ATGATCATTT 2220  
 25 CTTATATCTC AAACATAAAA TGAATTmymS TGAAGACGCT ATGATCAGGA TGACACGTAT 2280  
 TATAGAATCT TTCGATAGAG AAACGAATCA ACGTATCCGA CATCACGAAA TG 2332

30 (2) INFORMATION FOR SEQ ID NO: 478:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 865 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 478:

40 TTTACTACCC AGTATCTCTT TTTATAAATT ATATAGCCAC CACATATGGT GGAAAGTCTT 60  
 TTTAATTAGA ATTTTGTTTT TTCAGTTAAG AAAGCTTCTA ACTCTGAGAT TGGCATACGA 120  
 ACTTGTTCCTA TTGAATCTCT GTCACGTACT GTAACCTGAT TATCTTCTAA TGAATCAAAG 180  
 45 TCGAATGTtA CACAATAAGG TGTaCCGATT TCATCTTGAC GACGGTATCT TTTACCGATA 240  
 GATtGTGATT CATCGAAATC GATTGAGAAT TTAGAACTTA ATTGCTCAA AATCTTAATC 300  
 50 GtTCGCCAGA TAATTTCTTA CTTAAAGGTA AAATCGCTGC TTTATATGGT GCTAATGCAG 360  
 GATGGAAGTG TAAACTGTA CGTGCACTT TACTACCTTC AACGCCTTCT TCATCATATG 420

55

GAATATATTT TTCGTTTCGTT TCTGGATCAT GGTATCTGAA ATCTTCACCA GAGTGTTCAG 540  
 CATGTTTACG TAAGTCGAAG TCTGTACGAC TTGCGATACC CCATAACTCA CCCCCAACCA 600  
 5 ATGGAATTT ATATTCAATA TCAGTTGTTG CATTTGAGTA ATGAGATAAT TCATCTTCAT 660  
 CATGATCACG TAAACGCATA TTTTCACTGC TCATATTTAA GCTTGTTAAC CAGTCACTTG 720  
 CAAAAGTTTT CCAATAATTT TGCCATTGCA TTTCTTCTCC AGGTTTACAG AAGAATTCAA 780  
 10 GTTCATTG TTcAAATTCT CTTGTTCTGa AAATGaAGTT ACCTGGAGTG aTTTCaTTAC 840  
 GGaATGaTTT ACCAATTTGG ACCGG 865

## (2) INFORMATION FOR SEQ ID NO: 479:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1444 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 479:

25 TAATGAGTAA ATAAGTACCA GATAAAATCA TGAATATCAT CCAACATGCG GTTAACTCTA 60  
 CTAAATAATT AATAATAGTA TTTTCAGTAA ACAAGAAGT ATGTATACTT CGCATCACAT 120  
 TAGAATACGT ATGTTTCGCA TTTTGATCTG CAACAAATTG ATTGTTATGA TCTAGAAAGA 180  
 30 CGTAACGTTG ATTTCTCGCC ATATCACTCA GTGTAATTG TTTGTTATAT GGTTTCATCAA 240  
 GTATGCTAAC TTTACTTACA AAGAATCCTT CATATTGTTG TTCAACTTGA TGTACTGCAT 300  
 CATTTAATGT TTGATGCGTT TTTACATCAC TGTCAACAAA AAACCTCATTC TTATAAATAT 360  
 35 TATTTTCAAC TTCTGGAAAG AACAAGTAAC CAATGCCCGA AATGGTTAAA GTGATTAAAC 420  
 GTGGAGCAAT AAATATTGCT GCATAGAAAT GTAATCTTTG TAATGGATTA AATGTATTTT 480  
 TCATATTTCC CTCCCAATTG GCTATTATAC GGTGTCAATT CTGTGATGTG TGTGAACAAA 540  
 40 CTGTGACAAT ATTTATTTTC TAGAAAAATT TAACGATGAT TTGTGATTTT TAGAAAAATG 600  
 AACTTTTAAG TTGGAATGTT TGAAGAAAAT TGATTATTCG TATGTTTTAT CAAGCAGCTA 660  
 TGATAAAATT TAAACATAAT ACAATGCGAG CCATTTAACG ATCTATGTTT AAATGGACAT 720  
 45 CGATATTGTA TGAATTCGTT GTAACAAGCA AGCATTCTTA TGTGAACGAA CCAAAGGGGA 780  
 AAGTAACATG ATTAATAAAG AACAATTAGA TCTTTTATAT AAATTAAAAA AAGAAGTTGA 840  
 50 AAAGTCGCGA AATGAAGCAC TTTTACATAC AATTAAACCA GTAATTAAGA AAGTATATTT 900  
 GCAGCAATAT ACATGTTTCGT TCGTTGGACA TTTTCTGCA GGTAAATCGA CACTGATAAA 960

55

TATTGTGTCA GTTTCAGACA ATCACGATAT TATTGCTAAT TTGCCGAATC AAACGTATGC 1080  
 CAAATTATCT AATTATGATG AAGTAAGGGA AATGAATCGC CAAAATGTCTG ACGTTGAATC 1140  
 5 TGTAGAAATT AATTTTCAAT CAGCTAAATT TGAAAATGGG TTTACGTTGC AAGATACACC 1200  
 AGGTGTTGAT TCAAATGTTG CATCACATCA GTCAATAACA GAACAATATA TGTATACAAG 1260  
 TAATATGATA TTTTATACGG TTGACTATAA CCACGTTCAA TCTGAACTTA ACTTTAAGTT 1320  
 10 TATGAAGCAT ATAAATGATG TTGGAATACC TGTGTGTTT ATCATTAAATC AAATTGACAG 1380  
 CATCCAAGAC GATGGAATTG TCATTCTCTA CGTnTTAAAT CTCGAGTTGG AAAAATCAAT 1440  
 TGGC 1444

15 (2) INFORMATION FOR SEQ ID NO: 480:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6309 base pairs  
 (B) TYPE: nucleic acid  
 20 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 480:

GCAGCAGCTT ATCGTGAGTT ATCATTATTA TTACGTAGAC CTCCAGGTCG TGAACCATAC 60  
 CCAGGTGACG TATTCTACTT ACATAGTAGA tATTAGAAA GAGCAGCAAA ATTAAACGAT 120  
 30 GACTTAGGTG GCGGTTCAAT TACTGCATTA CCAATTATCG AAACACAAGC TGGTGATATT 180  
 TCAGCTTATG TACCAACAAA CGTTATTTCA ATTACAGATG GACAAATCTT CTTACAATCT 240  
 GATTTATTCT TCTCAGGTGT AAGACCAGCG ATTAATGCCG GACAATCTGT ATCTCGTGTT 300  
 35 GGTGGATCTG CACAAATTAA AGCAATGAAG AAAGTTGCTG GTACGTTACG TCTTGACTTA 360  
 GCGTCATACA GAGAACTTGA ATCATTTGCA CAATTCGGTT CAGACCTTGA TGAATTTACT 420  
 GCAAGTAAAT TAGAACGTGG TAAACGTACT GTTGAAGTCT TAAAACAAGA TCAAAACAAA 480  
 40 CCATTACCAG TCGAACACCA AGTGTGATT ATTTATGCAT TAACAAAAGG ATATTTAGAT 540  
 GATATTCCAG TTGTAGATAT CACACGTTTT GAAGACGAGT TAAACCACTG GGCAGAATCA 600  
 AATGCTACTG AACTGTTAAA TGAAATCAGA GAACTGGTG GCTTACCAGA TGCTGAGAAG 660  
 45 TTTGACACAG CAATTAACGA ATTCAAAAAA AGCTTTAGCA AATCTGAATA ATAAACAAGT 720  
 TTAGTATAAG GTGGTGAGAT AGTGGCTTCT CTAAAGAAA TAGATACTCG AATAAAATCA 780  
 50 ACCAAAAAAA TGAAGCAGAT TACGAAAGCG ATGAACATGG TATCAAGTTC AAAACTTCGT 840  
 AGAGCTGAAA AAAATACAAA ACAATTCACA CCATATATGG ATAAATGCA AGATGCAATT 900

55

	ACTAGAAGTG GATATTTAGT TATCACGAGT GATAAAGGTT TAGCAGGTGC ATATAGTGCA	1020
	AACGTGCTTA AAAAATTGAT TACTGATATT GAAGCGAAAC ATCAAGATAG TAGCGAATAC	1080
5	AGTATTGTAG TTTTAGGGCA ACAAGGTGTT GATTTCCTAA AAAATAGAGG TTATGACATT	1140
	GAGTATTCTC AAGTAGACGT ACCTGATCAA CCTTCTTTCA AATCTGTTCA AGCACTAGCT	1200
	AACCATGCTA TAGACTTATA CAGTGAAGAA GAAATTGATG AATTAAATAT ATACTATAGT	1260
10	CATTATGTCA GCGTTCCTGA AAACAAGCCT ACATCTAGAC AAGTATTACC ATTATCTCAA	1320
	GAGGATTCTA GTAAGGGGCA TGGTCATTTG TCTTCTTATG AATTTGAGCC AGATAAAGAA	1380
	TCTATCTTAA GTGTAATCTT GCCTCAATAT GTTGAGAGTT TGATTTACGG AACAATATTA	1440
15	GACGCAAAAG CAAGTGAGCA TGCAACACGT ATGACTGCGA TGAAAAATGC CACTGATAAT	1500
	GCAACTGAAC TTATTGATGA CTTATCATTGA GAATATAACA GAGCGAGACA AGCAGAAATT	1560
20	ACGCAACAAA TTAAGTAAAT TGTGGTGGT TCCGCAGCGC TTGAATAATA TTTAAAGGAG	1620
	GAAAATAGCA TGGGAATTGG CCGTGTAACCT CAAGTTATGG GTCCTGTAAT TGATGTTCTGA	1680
	TTTGAACATA ACGAAGTTCC TAAAATTAAT AACGCCTTGG TTATTGATGT GCCTAAAGAA	1740
25	GAAGGTACAA TACAATAAC ATTAGAAGTT GCGCTGCAAT TAGGTGACGA CGTTGTTCTG	1800
	ACAATTGCGA TGGATTCAAC TGATGGTGTC CAAAGAGGCA TGGATGTAAA AGATACAGGC	1860
	AAAGAAATTA GTGTACCTGT TGGTGACGAA ACATTAGGTC GTGTATTTAA TGTACTAGGT	1920
30	GAAACAATTG ACCTTAAAGA AGAAATTAGT GATTCTGTTT GCCGCGATCC TATCCATCGT	1980
	CAAGCACCAG CATTCGATGA ACTTTCAACA GAAGTTCAAA TTTTAGAAAC AGGTATTAAA	2040
	GTAGTAGATT TACTAGCACC TTATATTAAA GGTGGTAAAA TCGGATTGTT CGGTGGTGCC	2100
35	GGTGTAGGTA AAACAGTATT AATCCAAGAA TTAATTAACA ACATCGCTCA AGAGCACGGT	2160
	GGTATTTCTG TATTCGCCGG TGTAGGTGAA CGTACTCGTG AAGGTAACGA TTTATACTTC	2220
	GAAATGAGTG ACAGTGGTGT AATTAAGAAA ACAGCCATGG TATTCGGGCA AATGAATGAG	2280
40	CCACCTGGTG CACGTATGCG TGTTCGATTA TCTGGTTTAA CAATGGCTGA ATATTTCCGT	2340
	GACGAACAAG GTCAAGACGT ATTATTATTC ATCGATAACA TTTTCAGATT TACACAAGCT	2400
45	GGTTCTGAGG TATCTGCATT ATTAGGTCGT ATGCCTTCTG CAGTAGGTTA CCAACCAACA	2460
	CTTGCTACTG AAATGGGACA ATTACAAGAA CGTATTACGT CTACAACAAA AGGATCAGTT	2520
	ACTTCTATTC AAGCGGTATT CGTACCTGCC GATGACTATA CTGACCCAGC GCCTGCGACA	2580
50	GCGTTTGCCC ATTTAGATGC AACTACAAAC TTAGAACGTA AATTAAGTGA AATGGGTATT	2640
	TATCCAGCCG TGGATCCATT AGCGTCTACA TCAAGAGCAT TGGAACCATC AATTGTAGGT	2700

55

	CAAGATATCA TTGCTATCTT AGGTATGGAC GAATTATCTG ATGAAGATAA ACAAACAGTT	2820
	GAACGCGCAC GTAGAATTCA ATTCTTCTTA TCTCAAACT TCCACGTAGC GGAACAATTT	2880
5	ACTGGTCAAA AAGGTTCTTA TGTACCTGTT AAGACAACAG TTGCAAACCTT TAAAGATATC	2940
	TTAGATGGTA AATATGACCA TATTCCAGAA GATGCATTCC GTTTAGTTGG TAGCATGGAT	3000
	GATGTTATTG CAAAAGCTAA AGATATGGGT GTTGAAGTAT AACAATTAGG AGGAATGGAT	3060
10	AATGAATACA TTAAACCTAG ATATTGTCAC TCCTAATGGT TCTGTTTACA ATCGTGATAA	3120
	TGTTGAACTC GTTGTATGTC AAACAACAGC TGGTGAGATA GGTGTCATGA GTGGACATAT	3180
	TCCAACTGTA GCTGCTTTAA AACAGGCTT TGTAAGAGT AAATTTACG ATGGAACTGA	3240
15	ATATATTGCT GTAAGCGATG GCTTTGTTGA AGTTAGAAAA GATAAAGTTT CAATCATTGT	3300
	TCAGACTGCA GAAACTGCAA GAGAAATTGA TGTGAAAGA GCTAAATTAG CCAAAGCAAG	3360
20	AGCAGAGTCT CACTTGAAAA ATGATGACGA CAATACTGAT ATTCATAGAG CCGAAAGAGC	3420
	TTTAGAGAGA GCAAATAACC GTTTGCGTGT GGCTGAATTA AAATAGTAAA TAAAGGGTCG	3480
	AAGATGTGAT TTCATATCTT CGACCCTTTT TTGAATTATA TTGATTTAAA GATACAAAAC	3540
25	ATGAGAGGGG GGAAGGAATT GATAAAGAAC CATTAAAGAT TTATGATGTA GTGGTTCTTT	3600
	ATCATTAAAC ACAGCTAATG TGTATTTAAA AATAGGaayA CATgAGTAAA ACTCATGTAT	3660
	AAGAAATACT AATTTCTAAA GAAAAAGTAT TTCCTTATGT TGGGGCCCCG TCAACTACTG	3720
30	CCAAATACAA CACTATAGAG TCTAGACATT GATTTATGTC CGACTCCCAA GAATAGTTTT	3780
	ACTTTTTTAC AATCACTAAT AGATTGCTAA AATCAAAATT TCCTTCACCA CTATCTACAG	3840
	TCGACATTTT ATTTTTTGAA ATTATCTACA TTTTTTCATA CCAAGATATT TTATAGTTAT	3900
35	GATATTTATG TAAAAAGAAT TATATAGTAA GTTAGCTTAA ACTTTACTAA AAACGGGTAT	3960
	TAAACTTTGT ATCATTATTT AAATTTTTCA TGTACAATGT AATACAGTAA TCTTATGAGG	4020
40	TGATAAAATG GATTATATCG GACAATATGC AGTTATCCAT TTAGTGTTAC ATGTTGTATG	4080
	TATTTGTATT GCCTATTGGG CTTTACAATC AATTAGATTA GATCAATTTT TTAAAAAAGG	4140
	ATACGCCACT CAATTACAAG TGTGTATGAT ATTTGTTGCT ATTTTATTAG GCACTGCAGT	4200
45	AAGCAATTTT ATTGTAGATT TGTTACAATA CTCGACGCAG GTAAAATATT TAATAAAATA	4260
	AGTCTAACTC TATGATTTGT AATCAAACT AGATATAATT AAATAATGAC TTAAATAAAT	4320
	TTTAAATAG GGAAATGTAA AGTAATAGGA GTTCTAAGTG GAGGATTTAC GATGGATAAA	4380
50	ATAGTAATCA AAGGTGGAAA TAAATTAACG GGTGAAGTTA AAGTAGAAGG TGCTAAAAAT	4440
	GCAGTATTAC CAATATTGAC AGCATCTTTA TTAGCTTCTG ATAAACCGAG CAAATTAGTT	4500

55

	GACGTTACAT ACAAAAAGGA CGAAAATGCT GTTGTCGTTG ATGCAACAAA GACTCTAAAT	4620
	GAAGAGGCAC CATATGAATA TGTTAGTAAA ATGCGTGCAA GTATTTTAGT TATGGGmCCT	4680
5	CTTTTAGCAA GACTAGGACA TGCTATTGTT GCATTGCCTG GTGGTTGTGC AATTGGAAGT	4740
	AGACCGATTG AGCAACACAT TAAAGGTTTT GAAGCTTTAG GCGCAGAAAT TCATCTTGAA	4800
	AATGGTAATA TTTATGCTAA TGCTAAAGAT GGATTAAGG GTACATCAAT TCATTTAGAT	4860
10	TTTCCAAGTG TAGGAGCAAC ACAAATATT ATTATGGCAG CATCATTAGC TAAGGGTAAG	4920
	ACTTTAATTG AAAATGCAGC TAAAGAACCT GAAATTGTCG ATTTAGCAAA CTACATTAAT	4980
15	GAAATGGGTG GTAGAATTAC TGGTGCTGGT ACAGACACAA TTACAATCAA TGGTGTAGAA	5040
	TCATTACATG GTGTAGAACA TGCTATCATT CCAGATAGAA TTGAAGCAGG CACATTACTA	5100
	ATCGCTGGTG CTATAACGCG TGGTGATATT TTTGTACGTG GTGCAATCAA AGAACATATG	5160
20	GCGAGTTTAG TCTATAAACT AGAAGAAATG GCGGTGGAAT TGGACTATCA AGAAGATGGT	5220
	ATTCGTGTAC GTGCTGAAGG GGAATTACAA CCTGTAGACA TCAAACTCT ACCACATCCT	5280
	GGATTCCCGA CTGATATGCA ATCACAATG ATGGCATTGT TATTAACGGC AAATGGTCAT	5340
25	AAAGTCGTAA CCGAACTGT TTTTGAAAAC CGTTTATGC ATGTTGCAGA GTTCAAACGT	5400
	ATGAATGCTA ATATCAATGT AGAAGGTCGT aTGCTAAAC TTGAAGGTAA AAGTCAATTG	5460
	CAAGGTGCAC AAGTTAAAGC GACTGATTTA AGaGCAGCAG CCGCCTTAAT TTTAGCTGGA	5520
30	TTAGTTGCTG ATGGTAaAC AAGCGTACT GAATTAACGC ACCTAGATAG AGGCTATGTT	5580
	GACTTACACG GTAAATTGAA GCAATTAGGT GCAGACATTG AACGTATTAA CGATTAATTC	5640
35	AGTAAATTAA TATAATGGAG GATTTCAACC ATGGAAACAA TTTTGTATTA TAACCAAATT	5700
	AAACAAATTA TACCTCACAG ACAGCCATTT TTATTAATTG ATAAAGTAGT TGAATATGAA	5760
	GAAGGTCAAC GTTGTGTGGC TATTAAACAA GTATCAGGAA ACGAACCATT CTTTCAAGGG	5820
40	CATTTTCCTG AGTATGcGGT AATGCCAGGC GTATTAATTA CTGAAGCGTT AcTCAAACAG	5880
	GTGCGGTAGC TATTTTAAAT AGTGAAGAAA ATAAAGGTAA AATCGCTTTA TTTGCTGGTA	5940
	TTGATAAATG TCGTTTTTAA CGTCAAGTAG TACCTGGTGA TACTTTAACG TTGGAAGTAG	6000
45	AAATCACTAA AATTAAAGGA CCAATAGGTA AAGGTAATGC TAAAGCTACT GTCGATGGTC	6060
	AACTTGCTTG TAGTTGTGAA CTTACATTTG CAATTCAAGA TGTAATAATA AACAAAAAA	6120
	ACATTCAAAG ATTTAATGTG TTGGCATAAT CTTTGAATGT TTTTATTTT ACTCTTCTAA	6180
50	TTTTTCATCC TTAACTTTG GTTTAGACTG CaTCATTCGA TTAAATGATT TTTTAAATTC	6240
	TTCACCAGAT AATCCATCAT CAATAAGTTG GTTCTAATAA ACTTTCAGCA TACTGTTGGA	6300

55

## (2) INFORMATION FOR SEQ ID NO: 481:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1717 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 481:

	AGTTGCTACA CCAGACATGA TGGGTGAACT TGGTAAATTA GGTCTGTAT TAGGACCAAA	60
15	AGGTTTAATG CCAAACCTA AAAGTGAAC TGTAACAATG GATGTTAAA AAGCTGTTGA	120
	AGAAATCAAA GCTGGTAAAG TAGAATATCG TGCTGAAAA GCTGGTATCG TACATGCATC	180
	AATTGGTAAA GTTTCATTTA CTGATGAACA ATTAATtGAA AACyTCaATA CyTTACAAGA	240
20	TGTATTAGCT AAAGCTAAAC CATCATCTGC TAAAGGTACA TACTTCAAAT CTGTTGCTGT	300
	AACTACAACA ATGGGTCCTG GAGTTAAAAT TGATACTGCA AGTTTCAAAT AATAAATGAT	360
	ATAACAATT ACAGGCTGAA AGAAATATCT TTCAGTCTGT AAAAATATAT TGACAATAAG	420
25	TAATTTCCAA GTTATATTAC TTATTGTGAT TATTTTACCT AAGACAGTAG GAGTTATTTA	480
	TAACTTAAAA TTTATCCTGC CGAGGCTAAA ATTGACTTGA ACGTGATGAT CTATGATCTT	540
	TCAAGCACTT TTTGCCGTGG GTAGAAAGTG CTTTTTTTAT TAATTTTAAA AAAAGCACCA	600
30	AAAATTTAAA TGGAGGTGTC TGAATGTCTG CTATCATTGA AGCTAAAAAA CAACTAGTTG	660
	ATGAAATTGC TGAGGTACTA TCAAATTCAG TTTCAACAGT AATCGTTGAC TACCGTGGAT	720
	TAACAGTAGC TGAAGTTACT GACTTACGTT CACAATTACG TGAAGCTGGT GTTGAGTATA	780
35	AAGTATACAA AAACACTATG GTACGTCGTG CAGCTGAAAA AGCTGGTATC GAAGGCTTAG	840
	ATGAATTCTT AACAGGTCCT ACTGCTATTG CAACTTCAAG TGAAGATGCT GTAGCTGCAG	900
40	CGAAAGTAAT TTCTGGATTT GCTAAAGATC ATGAAGCATT AGAAATTAAA TCAGGCGTTA	960
	TGGAAGGCAA TGTTATTACA GCAGAAGAAG TTAAACTGT TGGTTCATTA CCTTCACACG	1020
	ATGGTCTTGT ATCTATGCTT TTATCAGTAT TACAAGCTCC TGTACGCAAC TTCGCTTATG	1080
45	CGGTTAAAGC TATTGGAGAA CAAAAAGAAG AAAACGCTGA ATAATTTTTA GCGTAAAAAA	1140
	ATTAAAAATA ATGGAGGAAT TATAAAATGG CTAATCATGA ACAAATCATT GAAGCGATTA	1200
	AAGAAATGTC AGTATTAGAA TTAACGACT TAGTAAAGC AATTGAAGAA GAATTGGTG	1260
50	TAcTGcAGCT GCTCCAGTAG CAGTAGCAGG TGCAGCTGGT GCGCTGACG CTGCAGCAGA	1320
	AAAACTGAA TTTGACGTTG AGTTAACTTC AGCTGGTTCA TCTAAATCA AAGTTGTAA	1380

TCCTAAAGTA ATCAAAGAAG CTTTACCTAA AGAAGAAGCT GAAAACTTA AAGAACAATT 1500  
 AGAAGAAGTT GGAGCTACTG TAGAATTAAA ATAATTCAAG TATCTTAAAC TTAATAATCA 1560  
 5 AAGTTTTATA GCAAGTATTG CTATAATATA ATGATTCTTT GAGAAGTTAA AACCCCGTTA 1620  
 TTTTGATAAC GGGgTtTTAT TCaTTTAAAG ACTGAGTGAA ATGTTATAAT TATAATGACG 1680  
 AGTTACAAAG TGAAGATGAG GTGGGAATAA TGAGTCA 1717

10

(2) INFORMATION FOR SEQ ID NO: 482:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1279 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 482:

GTAAATCTGT TACTCGAAAT GTAACGATTA AAGAAAAGGG CTCATCTCAA ACATATATTT 60  
 TGTTAGGCTA TCCAACAAAA GCACAGAAGA ATAGTCATAG CAAATATAGT GGAGTCTTTA 120  
 25 TATATAAGA CTTGAAATCA ATCGAAGATA CAAATAATGC TATTACGATT ATCACCATAA 180  
 TTACGGCTGT TATTTTCTTA ACAATTACAA CAGTCTTTGC GTTTTTCTTA TCGTCAAGAA 240  
 TTACAAAACC TTTAAGACGT TTAAGAGACC AAGCTACACG TGTATCTGAA GGGGATTACT 300  
 30 CTTATAAACC TTCTGTCACA ACGAAAGATG AAATGGGTCA ATTATCGCAG GCATTTAATC 360  
 AGATGAGTAC AGAAATCGAA GAGCATGTCTG ACGCATTATC CACATCTAAA AATATTAGAG 420  
 ACAGCTTAAT TAACTCTATG GTAGAAGGTG TCCTAGGTAT TAATGAGAGT CGACaAATTA 480  
 35 TCTTATCTAA TAAGATGGCG AATGATATTA TGGACAATAT TGATGAAGAT GCTAAAGCTT 540  
 TCTTATTAAG ACAAATAGAA GATACTTTTA AATCAAAACA AACTGAAATG CGCGATTTAG 600  
 AAaTGAATGC ACGATTCTTT GTTGTGACCA CAAGCTATAT CGACAAGATT GAACAGGGAG 660  
 40 GTAAaAGTGG TGTGTGTGTG ACAGTTCGTG ATATGACTAA TGAGCACAAT CTAGATCAAA 720  
 TGAAGAAAGA TTTcATTGCT AATGTATCAC ATGAATTACG TACACCGATA TCATTACTTC 780  
 45 AAGGTTATAC TGAATCAATT GTAGATGGTA TTGTTACAGA ACCGGATGAA ATAAAAGAAT 840  
 CGCTTGCCAT TGTcCTTGAT GAATCGAAAC GTTTAAATCG TTTAGTTAAT GAATTGTTAA 900  
 A1GTCGCACG CATGGATGCT GAAGGGTTAT CCGTAAATAA AGAAGTTCAG CCTATTGCAG 960  
 50 CGTTACTAGA TAAGATGAAA ATTAAGTATC GCCAACAAGC TGATGATTTA GGTCTAAATA 1020  
 TGACTTTTAA TTaYTGTAAG AAGCGTGTTT GGAGTTATGA TAaGGATCGC ATGGACCAAG 1080

55



TTACTTGTGA TGAAATGAA AGCGAAGATA TTTTATACAT TAAAGATACA GGTACAGGCA 1200  
 TTGCACCAGA ACATTTCACmA CAAGTATTTg ATCGTTTTTA TAAAGTTGAT GCAGCGnAnA 1260  
 5 ACCCCGnGGT AACCAmGTA 1279

(2) INFORMATION FOR SEQ ID NO: 483:

(i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 1144 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 483:

GAGCTGTTGT TACTTTGATG CCTGCAGCTT TATTACGGCT GACTTGGTAA TGATAAGTTT 60  
 20 CAGCATATTG CTCAATATAT GCTATATCAT ATTGAATGGT ACGAGGTGAT ACACCAAGTT 120  
 GATTAGCAAT GGTATTGATT GGAATAAACG TTTGCTCATG AATTAAAAGA TACAAAATTT 180  
 CGATTTGTCT ATAACCTAAC AACGTAATAT CCTCCTATTT GTAATTGTAA GCGATTCTT 240  
 25 AAAAACGTAG ATATGCAATC TCTTTCATAT TTTAATCCGA AAAATTGCAT ATCAAAATGT 300  
 TTATGGCGCA AGATTTTATA GGAACTTTAA AAATAAATTA rATATTCATG TTGACAATTT 360  
 AAAAATGTCG CAGTATATTT AGTTAGACAT CTAACGAAAT GGTGGTGCAA TAAATGGAAT 420  
 30 TCACTTATTC GTATTTATTT AGAATGATTA GTCATGAGAT GAAACAAAAG GCTGATCAAA 480  
 AGTTAGAGCA ATTTGATATT ACAAATGAGC AAGGTCATAC GTTAGGTTAT CTTTATGCAC 540  
 ATCAACAAGA TGGACTGACA CAAAATGATa TTGcTAAAGC ATTACAACGA ACAGGTCCAA 600  
 35 CTGTCAGTAA TTTATTAAGG AACCTTGAAC GTAAAAAGCT GATCTATCGC TATGTCGATG 660  
 CACAAGATAC GAGAAGAAAG AATATAGGGC TGA CTACCTC TGGGATTAAA CTCGTAGAAG 720  
 40 CATTCACTTC GATATTTGAT GAAATGGAAC AAACACTCGT ATCGCAGTTA TCTGAAGAAG 780  
 AAAATGAACA AATGAAAGCA AACTTAACTA AAATGTTATC TAGTTTACAA TAAATGATAA 840  
 GTGTGACTGG TAGAAATCAG TCACTTTGTC TTTAATATTA TAGTTAGATA TCTAATTGTT 900  
 45 AGTAAGCTAA TTATTGGAAA AGACAAGGAG TATTGAACAA TGAAAGACGA ACAATTATAT 960  
 TATTTTGAGA AATCGCCAGT ATTTAAAGCG ATGATGCATT TCTCATTGCC AATGATGATA 1020  
 GGGACTTTAT TAAGCGTTAT TTATGGCATA TTAAATATTT ACTTTATAGG ATTTtTyAGAm 1080  
 50 GAYAGCCACA TGATTTCTGC tAatCTCTCT AACACTGCCA GTATTTGCTA TCTTAATGGG 1140  
 GTTA 1144

55

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1158 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 484:

10 ACAACCAATT TTAATAAACA TGGTrTCAT gCATTTTCAA TATTTAGATG TCGACCGCAT 60  
 AATCGAAGAA TCGCCGACAA TAGTACTTAT CGATGAGTTA GCACATACGA ATATTTCTAG 120  
 15 AGATCGTCAT GAGAAACGAT ATATGGATAT TGAAGAAATT TTAAATCATG GTATCGATGT 180  
 TCATACCACT TTGaACATTC aaCATATTGA AAGTTTAAGT AGTCAAATTG AACTGATGAC 240  
 CGGTGTACAT GTTAAAGAAC GTGTACCCGA CTATTTTATA ATGAGCGCCG ATGTATTAGA 300  
 20 AGTCGTAGAT ATCTCACCTG AACAAATTAAT TAAACGCTTA AAAGCTGGCA AGGTATATaA 360  
 AAAGGATAGG CTAGATGTAG CATTTAGTAA TTTCTTTACG TATGCCACC TAAGCGAaTG 420  
 CGTACATTGA CGTTAAGAAC AGTTGCCGAC TTGATGAGTG ATAAAGAAAA AGTCCGACAC 480  
 25 AACCATAAAA CGTCACTCAA ACCTCATATT GCTGTGGCAA TTAGTGGGAG CATTTATAAT 540  
 GAAGCAGTAA TTAAAGAGGC ATTCCATATT GCTCAAAAAG AACATGCGAa GTTCACTGCT 600  
 ATTTATATAG ATGTATTCGA AAAAAACAGG CAATATAAAG ATAGTCAAAA GCAAGTGCAT 660  
 30 CAACATCTCA TGCTTGCAAA ATCATTAGGA GCAAAAGTAA AAGTAGTTTA TAGCCAAACC 720  
 GTTGCATTAG GATTAGACGA ATGGTGTAAG AATCAAGATG TAACCAAATT AATTATCGGA 780  
 CAACATATTA GAAATAAGTG GCGAGACTTT TTCAATACAC CTTTAATTGA CCATTTAATG 840  
 35 TCCTTTGAAC ATAGCTATAA AATCGAAATC GTTCCAATCA AACAAATACC TGTGAATTG 900  
 AAAATGAACA AATCACCCCTA TCGTCCTAAA GGCAAACGTT TCGCCATAGA TATGTTAAAA 960  
 40 ATGATTTTGA TTCAAATAAT TTGTGTAATG ATGGGACTGT GGATTTATCA ACTTGATAAG 1020  
 CATGAGTCTA GTACGATTAT TTTAATGATT TTTCTCATCG GCATCATTTT ATTATCCATT 1080  
 TGGACGCGGT CCTTCATCAT TGGCTTTTaG CAGCAATTat TAACGTATTT GTgTkTAATT 1140  
 45 ATkTTTTtAC GGAACCTA 1158

## (2) INFORMATION FOR SEQ ID NO: 485:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2224 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

EP 0 786 519 A2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 485:

	ATCATATGGT CGATTTAACA GATCCAACGT ACTGCTAAAT AATTACATGA CGTTTAAACA	60
5	TAGCATTGAT TATAACTATT TCTAAGTCTT CGCATTATTT GCGATGATGT GGAATAGTT	120
	ATTTTATTTT AAAAATATAA AAAAATAGAT GCAGCAAAAT TTTAAAGCAT TTTATTTTGA	180
	ACATATTAAA AGGGAGCGTA TCATAATGGA ATGTAATGTT TATATCGTAT GCATTACGGA	240
10	TAAATAATAT ATAAATCATT CTTGAGGAGT GAAAGAATAA TGAGAGACTA CACAAAGCAA	300
	TACATTAATG GCGAATGGGT AGAAAGTAAT AGTAATGAAA CGATAGAAGT TATAAATCCA	360
	GCAACCGAAG AAGTAATCGG GAAAGTTGCT AAAGGTAATA AAGCTGATGT TGATAAAGCC	420
15	GTGAGGCGG CAGACGATGT TTATTTAGAG TTCCGTCATA CATCTGTGAA AGAAAGACAA	480
	GCGTTATTAG ATAAAATTGT AAAAGAmTAT GAAAACAGAA AAGACGATAT TGTACAAGCT	540
	ATTACGGATG AATTAGGTGC TCCTTTATCA TTATCTGAGC GTGTCCATTA TCAAATGGGA	600
20	CTAAACCATT TTGTTGCAGC GAGAGACGCA TTAGATAACT ACGAATTTGA AGAACGCCGC	660
	GGAGATGATT TAGTTGTAA AGAAGCAATC GGTGTATCTG GATTAAATTAC ACCGTGGAAC	720
25	TTCCCTACAA ACCAAACATC ATTAAAATTA GCAGCAGCAT TTGCGGCTGG TAGTCCAGTT	780
	GTACTIONAAC CATCTGAAGA AACACCATTT GCAGCTGTTA TTTTAGCTGA GATTTTGTAT	840
	AAAGTCGGTG TTCCTAAAGG TGTATTTAAC CTTGTTAATG GTGATGGTGC TGGTGTGGG	900
30	AATCCTTTAT CTGAACATCC TAAAGTACGC ATGATGTCAT TTACAGGATC AGGCCCTACT	960
	GGTTCTAAAA TTATGGAAAA AGCCGCTAAA GATTTTAAAA AGGTATCATT AGAGCTTGGT	1020
	GGCAAATCAC CATATATCGT CCTAGATGAC GTAGATATTA AAGAAGCGGc TAAAGCAACa	1080
35	aCAGGCAAAG TTGTTAATAA TACTGGTCAA GTATGTACAG CTGGTACACG TGTTTTAGTG	1140
	CCTAACAAAA TTAAAGATGC ATTCTTAGCT GAATTAAAAG AACAAATTAG CCAAGTGGT	1200
	GTGCGTAATC CAAGAGAAGA TGGTACACAA GTAGGCCCTA TCATTAGTAA AAAACAATTT	1260
40	GATCAAGTAC AAAATTATAT TAATAAAGGT ATTGAAGAAG GTGCTGAATT ATTTTATGGT	1320
	GGTCCTGGTA AACCAGAAGG ACTTGAAAAA GGATACTTTG CACGTCCGAC AATTTTATT	1380
	AATGTAGATA ATCAAATGAC GATAGCACAA GAWGAAATTT TTGGGCCAGT AATGTCAGTT	1440
45	ATCACTTATA ACGATTTAGA TGAAGCGATT CAAATTGCAA ATGATACAAA ATATGGTTTG	1500
	GCAGGATATG TTATTGGTAA GGACAAAGAA ACATTGCATA AAGTAGCTCG TTCTATTGAA	1560
50	GCAGGTACAG TAGAAATAAA CGAAGCAGGT AGAAAGCCAG ATTTACCATT TGGTGGCTAT	1620
	AAACAATCTG GTTTAGGTCG TGAATGGGGC GATTATGGTA TTGAAGAGTT CTTAGAAGTG	1680

55

AGTGCACATG ACTAATTAAG TTTTGTGTAC TGTTTTAATT TTGCAATTTT TATAAATAGA 1800  
 TTTTGTAAAT AAAATAAAAA TTTGCTATAG TTATTCATGT ATTTAAAAGG TTGGGGATTA 1860  
 5 GCATAATGGG ATTGTGCTAG CACAGTTATT TATGCATTGT CATGCCTATC TATTACTTAC 1920  
 TAACTAAAAA ATAATGAAAT GGGTGTAAAC TATATGCCTG AAAGAGAACG TACATCTCCT 1980  
 CAGTATGAAT CATTCACGA ATTGTACAAG AACTATACTA CCAAGGAACT CACTCAAAAA 2040  
 10 GCTAAAACTC TTAAGTTGAC GAACTATAGT AAATTAAATn AAAAAGAACT TGTTCCTAGCT 2100  
 ATTATGGAAG CACAAATGGA nAAAGATGGT AACTATTATA TGAAGGTAT CTTAGATGAT 2160  
 ATACAACCAG ATGGTTATGG TTTTTTAAGA ACAGTGAACCT ATTCTAAAGG GGAAAAAGAT 2220  
 15 ATTT 2224

(2) INFORMATION FOR SEQ ID NO: 486:

(i) SEQUENCE CHARACTERISTICS:  
 20 (A) LENGTH: 1690 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 486:

ACATTACnTT GAATCGAAAG TTTCATAAAC GACTTGAATG CCAGTTTCTT TTTCAAATTT 60  
 30 CTTAATTAAC TCTGGATCAA TATATTCGCC CCAATTGTAT ACGTAAATTT TTTGATTTGT 120  
 ATGCACTTGT TCTTTAGATT TAAACCAATG ACTTAAAGTA AGACAAAGCA TACCCACAAC 180  
 TAATGCACCT ATAATGAGTT GTAAAAATCG TTTCAATTATT TTACACCTCG CTTGATTAGT 240  
 35 TTTTCTTAT TTATCACGTA TTGAATCAAA TAATATCCTA GTATTCCTAA TACAATAACA 300  
 GCAAACAATA ATGTTGAAAT CGCATTAATT TCCATACTAA TTCCTTTTCT CGCCATAGCA 360  
 TAAACTTCAA CTGATAACAC ACTAAAGCCA TTACCAGTAA CGAAGAACT TACTGTGAAA 420  
 40 TCGTCTAGTG AATAAGTTAA AGCCATAAAG AATCCTCCTA TAATAGAAGG TAAAATATTA 480  
 GGAATAATAA TGTTGCTTAA TAATTGTGGT TCAGTCGCTC CTAAATCTCT TGCAGCATTT 540  
 AACATATTAT TATTCATyTC ATACAGTTGT GGTAAGACGA TAATCACAAC TATAGGTATG 600  
 45 CAAAATGCAA TATGAGATAT TAGAACTGTC CaAAAKCCTA AACCAAGACC AGTAAAATGG 660  
 CCAATCGTTG TAAACATAAT TAAGAATGAT GCACCTATGA CAACGTCGGA TGATACCATC 720  
 AAGACATTAT TCAATGTTAG TAAAGTTACT TTAAACTTTT TATTTCTTAA ATAATAAATA 780  
 50 GCAATGGCAC CAAATGTACC AATAACTGTA GAAATTGAGG CTGCTAAAAG TGCTACAGCT 840

55

AATGTAAAAT GTTCAAAGTG AATCATATTA CCAGCCGAAT TGAATGAATA GAACATTAAA 960  
 AAGAATATTG GGATGTATAA AATCGCTAAA AGTATCCCGA TATACAGCTT TCCATACCAT 1020  
 5 TTCATATGAT TCACCCTCTC CCATTAGATG ATTTTGTAAT GATTAAAATG AATGCCATAA 1080  
 ATACAATTAA GAATATAGCT ATAGTTGATC CCATACCATA ATTTTGAATT GTTAAAAATT 1140  
 GTTCCTCTAT TGCCGTACCT ATATTTATGA CTTTATTACC TGCAATTAAT CTTGTAATCA 1200  
 10 TAAATAATGA AAGTGATGGA ATAAAGGTTA CTTGAATCCC AGTCATAACA CCTTCTTTTG 1260  
 TTAACGGCAT GATTACTTTT CTAAAAGTAT AGAAAGGACT GGCACCTAAA TCACTTGAGG 1320  
 CCTGCAATAA ATTATTAGGA ATTGCTTTCA TGCTATTAAA TATAGGTAAA ATCATAAATG 1380  
 15 GTATATAAAT GTAACCTGCC ACTACTAAAA ACGCACCAGT TGTAAATAAC AAATTGAATG 1440  
 ATGGTAAATT AAATAAGTGG AAAATTGATT AATCAGCCA TCATGACTTA ATAAACCTAT 1500  
 AAAAGCATAT GTCTTTAACA ATAAATTTAT CCATGTTGGA ATAATCATT TCAATTAATA 1560  
 20 GATATTTTGA AATTTGGAAC GAGTAATATA ATAGGCAGnT GGATAACTGA TAGTCAAGGT 1620  
 AATAATTGTT ATTGAAGCGG CATATAAAAT TGAATATGCA AACATTTTCA AATATTTTGT 1680  
 25 AGTAAAAATT 1690

## (2) INFORMATION FOR SEQ ID NO: 487:

## (i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 2112 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 487:

35 ACGAAAAGAA ATATTATGAT GAACAAAAG AAAGAATAAC GATTTATATG AAGTACAATG 60  
 TGAAAGGTTA TAAAAATATA AGCTTCGCTA ATTTTAAAGA AAACCCAATG GATGGTTATT 120  
 40 CTATTAGTGG TTATATAAAT AAtGaTAAAA AGTTATCATT TACAGCTGGT ATAAGATCTG 180  
 TTGATGATTT TCAATTTGAT ACCGATATTT CTTATACAGA TGAATTGGGT AGAAAATTTA 240  
 ATAAAAATCC TAAGTCAGTT TCTGAAATAA AAAAAGAGCA AAATACGTCC AATAAATAAT 300  
 45 TGTTCATATT GTGATGAAAC AAAAATATAA GTCATTAGAT GAGTTTAACT ATGTTATAAA 360  
 TATTTGTAGT ATCTATAAAA ATCTCGACAC TATTAAAATG ATAAAGTGCC GAGGTTTCT 420  
 50 TACTTATTTA GTTAATTCAA AGTTTATGCC AGATTCATAA GAATTTGTGA CACTTTTAAT 480  
 AGTGTAcCAT TGATTATTAC AATTTATCAA ATGGTCCTTT AGAAGGTATA AATAACAAAA 540

55

	TATTATTATG TTCAAACTT TACGCTCCAA AAAGTAAAAA GGAAGTTAAG CAATGTTTAG	660
	TTGCTTAaCT TCGGaTATTG AACGCATCAG TCCAATTTGA CATAGAGCCT TTTTtagTTC	720
5	TTGATGTTTC TCTTTAAaAC CTTGCATATT TTACAAAAAG AAAAATTAGC AGTATAATTA	780
	AGACAACGAA AATAAGTATT TACTTATACA CCAATCCCCT CACTATTTGC GGTAGTGAGG	840
	GGATTTTTAT TGGTGCGGCT ATATGTCACC TATTTGTAT TCGTCTACT TAGCCAATAA	900
10	GAAAAAAACG CAATGGCACA GCCACTGATG ACTGGTGCTA TGATGTGAAC GaAAATAAGC	960
	ATCACCTTAT ACACCTCCTC TCTGCGTCTA AATTGACGSc TGAGaGrTAG GcGACTCTAC	1020
	TATTATATCA TCGGCAAATA TACAAGCACA GTCAC TTGCT TCTGATAAGT TATATGATTC	1080
15	TAGCTGATAG ATTGAATCGT CTACACTTAA TTGGACAAAT TCTATGAGAA TAGATATTGT	1140
	TAATTTAAGA AAGTAGGCGA TTTTATTATG ACAAGAGAAA GAAGATCATT TAGTTCAGAG	1200
	TTTAAGTTAC AAATGGTTAG ATTATATAAA AATGGTAAGC CTAGGAATGA AATTATACGC	1260
20	GAGTATGATT TCACACCTTC GACGTTTGTA AATGGCGGTT ATAAAATGTA GGAAAATGGA	1320
	TAAAGCAACA TCAAAACACG GGTACATTCA ATCACCAAGA TAACTTATCG GATGAAGAAA	1380
25	AAGAGCTGAT TAAATTACGC AAAGAAGTTC AACATTTAAA AATGGAGAAC GATATTTTAA	1440
	AGCAAGTAGC GCTGATTATG GGGCAAAAAT AGAAGTCATT CAAAAGAATG CACATCAATA	1500
	TTCAGTATCA GCAATGTGTA AAGTCCTGAT AATACTAAGA AGTACCTATT ATGATTCTAT	1560
30	AAAAAGAAAA GATAATAAAA TCACTAAAGA TGATTCAAAC ATAGAACATG CCGTCATAAA	1620
	TATTTTAAAT TCTAATAGAA AAGTCTTTGG TACAAGACGA ATTAAAAATC ATTTAAATGA	1680
	CAAGGGTCTC ACTGTATCTG GACAAAAGAT AGGTCGATCA TGAAAAATC TAGTTTCTGT	1740
35	TTATACGAAA TCTAAATACA AAAATCATCT AAAAGAACT AATGAAAAC GAATTAAAAA	1800
	TCTTTATTAT TAGCTGCTGG TGTATTATT GTTAGTCCAA TTTCATTATC TTATAATTCA	1860
	GATGTAGCTC ATGCTGAAGA TAAGTTAGAC CATCTCAAG CAAAGGTAAT ATATTTGAGT	1920
40	AACCAAAATT TATTTGATGA ACTTGAGAAA AAAGGTTATA AACTGGAAGA TATATTTACA	1980
	AAAGAAGAAA TAAAAAATA TAAAGCTGAA GACCAATTGA GAGCGGGTAA AACTCAATAT	2040
	GTAGAAACAG GTAAAGATAC TGCAACATTA TATCTTTCTT CTGCATATAC AAAACAATA	2100
45	GCTGCTTTAG GT	2112

(2) INFORMATION FOR SEQ ID NO: 488:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 454 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 488:

5 GTAGGCACAC ATCTGCCATA TAAACATTCT TTTATACTAG TGTTCATATA TAGTGTAGAG 60  
 TTATAGTCTC CTTCTTGAAT CTCGAATAAT TCAATCAACC TATCAACCTT AGTCTCTTCC 120  
 GTTACTTCTT TTTCAATATC AACTATGAAG GGGATATCAA TTGGAATAAA ACTTGACGTC 180  
 10 GAACACTTAT TTGTATTTGG ATGAAAACGA ACGAATCCAT CACTAAATCC TGTTGAAAAA 240  
 AATATTTTTC CTTGTGATAG ATCCGGATTT TCTCGCGCCC ATTTAATTAA TTCATCTAAT 300  
 CTCATTTCTT TTTTAACTTT GATTTTCATT GTTATATCTC CTCTTGAACA GTAAATTTAT 360  
 15 CGTTAACTGA TACGTATCCA GTCACATTAC ATAAGATGCT ATCAACATCA AAAGTCACAC 420  
 AACAGTTGCG TTCAACATCA TTGAATAGA ATCT 454

## (2) INFORMATION FOR SEQ ID NO: 489:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1372 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 489:

TTGTCAGAAT TAGAATGCTT TTGAGTTACT TCATAATACT CATCAGTTTT TTGTGTATCC 60  
 30 TTTTGACTTT TATTTATTTT TTTCCACTTA CCAGTATGAC TTTCTTTTTT TACAGTTATT 120  
 TTCGGTTTGT TAAAAAATG ATAACCGATG TTCTTTTTTG TATCTGTGGA CTCTAAAAAG 180  
 ACTGAATTGT TTTCTGATT ATCAGAATTG GTTGTGTTGT TATCGTCTGT ATATAACGTA 240  
 35 TACCCATTCT CTTTGGCATT TTCAATCGTT GTAACGGAT TCTTTGATGA ATCAGTACTT 300  
 TTAATGCCAG TTCCTAAGAA GACAATTTTA TCGTTTAATA TGAAATATGA TTTTTTGGCA 360  
 GTTAAAGTTT TGTCTGATT TTCAAAATCC ATTCCGATAC TAGCATGTTG GTCATCAACT 420  
 40 TTTGTTCCGC CAACAAAAGT TTTACTCGAC TTTTATCAT CCGTATCTTT TAATATTTCA 480  
 TTGTCTAAAG TTGTTGTACC TGATAAACGT TTCATATCGG CTGTCACCCA GAAGTTATCA 540  
 TGATAGTGTT TGACATCGCT GTTATATAAA TAAGACATTC CAGCACCAGT GTGCCAACCT 600  
 45 TTTAAATTCT CTCCGTTGAT ACTTTCATAG CGTGCTACGT TTTTCGACGT CATACTTAAA 660  
 CCAATGCAA AGTCTAAGTC TTTGTTATGA TAGGTGACAC GATCCATGTC ATTATATATT 720  
 50 TTAAGTTGTT GTGTTAATCC GTTTTATAGAA ATACTGTTAT CTGTCATTAA AGACTTCATT 780

55

ACTGAAGATT TGACAATCTT TTTATACTTA GCTTTTGTG AATCATCCAT GGCATCACTT 900  
 AATCTCAACA ATGATTTTCAT TACTGTTGCA GATGCTGAGT GACTGGTTTC ATTTTCACGA 960  
 5 CTGATAGCTC TACCTCGTGA TAAATCCATC ATTTACACCTT TATAAATGAG TGGCATAAAT 1020  
 CCGTCGTCAA TCCATGACTT TAAGGTTGTA TCATTTTGGG TTTTATCATT AAAAGGTGTT 1080  
 TCTTTTATCA TCGGCATCAT TTGAGAAATC CCTCTAAGA GTACAACGCC ATAAGCACCA 1140  
 10 GTGTATGGAA CGTCTTGATG ATCAATGTAA GAGCCATCTT TATAAAATCC ATTACGTTCT 1200  
 TTACCACTGG CAGAATCTTG AACGTAAGTG AAGACTTTAT TAAATGAATC TATAGACTTT 1260  
 TTCATCATAT CTTTATCTTC TTCGATAATA CATTTCTAAA GTTTCACCTT AGAAATGTCT 1320  
 15 ACTAnATTTC CGCCTTTAGC AAGTTCAGnT TTTCTACAC AAGATAATAT TT 1372

(2) INFORMATION FOR SEQ ID NO: 490:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 564 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 490:

ACAACAATAT AGTAAAGCGT CGTTAATTAA ACAATGGGAC CAATTTGTTC GTCTTATATA 60  
 30 AATGTACTTT aCCTTTATTT TTACAAAAAT AGCATTTTCC TATGTCATTT AACTAAACAT 120  
 GTAAGTTCGT ATGAACGAGG TTTGTTAAAT AGATGATTCT AGGAAAATGC TTTTTCCTTT 180  
 TGACTTAGTT TAAATATTT TGCCACTTTG TACTGATAGT AGTTGCATTG TACTGTTGTG 240  
 35 CAGATTCTAT GCTATTAATT GAAACTGTT GCAATTTTGG AGTATTATTT AGTAATTGGT 300  
 CGACCTTTTC AACCATTGTA TTGATGTCAC CTTGAGGTAC TAAATAGCCA TTAAATCCAT 360  
 CTTGaATCAG TTCTGaTGGa CCATAATCTA CATCATAACT GATCACTGGT GTACCTACTG 420  
 40 AAAgCGACTC TAAATTGCT AAGCCAAAAC CTTCCATTTT ACTTGTCGAT AACATCAGTT 480  
 CTGCTTTAGC AATCTCTTCA TTAATATGCG TCTTAAAACC ATGAAATTTA ACATGTTCCA 540  
 GATATnATGA TAATCTTCTA CAAG 564

45 (2) INFORMATION FOR SEQ ID NO: 491:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1277 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 491:

5 TATCCACCCC ACGAAAGCCC CGGAAACTTA TTGTGTTACA AGATATATAA GCAGAAACGA 60  
 ACAACAGTTA ACAAATAAAA TGAAATTAAA CGTTTTAAAA ATGAAACAAA TGAAATCATC 120  
 TATTAGGTTA TGAAACTGTT TATAGCTTGA ATAGAAGCAT TTATTTTTTA GGAGGACAAT 180  
 TATTATGCGT CAAACATTTA TGGCAAATGA ATCAAACATT GAGCGCAAAT GGTATGTTAT 240  
 10 CGATGCTGAA GGCCAAACAT TAGGTCGTTT ATCATCAGAA GTAGCATCTA TCTTACGCGG 300  
 TAAAAATAAA GTAACCTTACA CACCACACGT TGATACTGGT GATTATGTAA TCGTTATTAA 360  
 TGCATCAAAA ATCGAATTTA CTGGTAACAA AGAACTGAC AAAGTTTACT ACCGTCACTC 420  
 15 AAATCACCCA GGTGGTATCA AATCAATCAC TGCTGGTGAA TTAAGAAGAA CTAACCCAGA 480  
 ACGTTTAATT GAAACTCAA TTAAAGGTAT GTTACCAAGC ACTCGTTTAG GCGAAAAACA 540  
 AGGTAAAAAA TTATTTGTAT ATGGTGGCGC TGAACATCCA CACGCTGCAC AACACCAGA 600  
 20 AACTACGAA TTACGTGGTT AATTAGAAGG AGGAAATGAC TTTGGCACA GTTGAATATA 660  
 GAGGCACAGG CCGTCGTAAA AACTCwGtAG CACGTGTACG TTTaGTACCa GGTGAAGGTA 720  
 ACATCACAGT TAATAACCGT GACGTACGCG AATACTTACC ATTGGAATCA TTAATTTTAG 780  
 25 ACTTAAACCA ACCATTGAT GTAACGTAAa CTAAAGGTAA CTATGATGTT TTAGTTAACG 840  
 TTCATGGTGG TGnTTCACCTG GACAAGCTCA AGCTATCCGT CACGGAATCG CTCGTGCATT 900  
 30 ATTAGAAGCA GATCCTGAAT ACAGAGGTTT TTTAAACGC GCTGGATTAC TTAATCTGTA 960  
 CCCACGTATG AAAGAACATA AAAAACCAGG TCTTAAAGCA GCTCGTCGTT CACCTCAATT 1020  
 CTCAAAACGT TAATTGTCGG ACGATATATA CAAAACACCT CGATATTATG TCGAGGTGTT 1080  
 35 TTTTGGCGT TTTTGGCGG AATATGGAAT GTGTAGAATA TAAATGAATT TTTACCTTCC 1140  
 CACCATAAAA GATGAAGAAC CATGAATGTG GAGAACAATA AATAGTTGGA TATTCTGTTA 1200  
 TTTTTTTGGA AGTGAAGTG GATTGGAAT ACTTTACTCn AAACGATTAA AAGGTTTAAA 1260  
 40 AAAACAACAA AnAGAAA 1277

## (2) INFORMATION FOR SEQ ID NO: 492:

## (i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 673 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 492:

TTATTGACAT TGT TTTTATC CAAAATTCAT TGTTAAGACA TTTTCTTTAT GAAATAATAA 120  
 TnATATTGAA GTATATTTTT ATTATTATTA AAAATAAATA AGGGGATACT TATGAGCACA 180  
 5 AATCAAACAT TTTTAATAIT TGTTATAGCA ATTATTCTAC TTACATCTGT AATAGGAATT 240  
 GTTGGACGAT ACATGAGTCG TCAACGTCTA TTAAAATCTA TGGAAACATT ATGGCAAACG 300  
 ATTTCTCCAT TAGAAGCTTT TATCAGACCG AACTCACATT TCGACTATGA GTATAAGCTC 360  
 10 TACAAGGAAA AATTTGAATC ACATTCATTA GTTGATGATA AAAC TTGGTC CGACTTAAAT 420  
 ATGAATGCAA TCTTTCATAA GATGAATTAT AATTTAACAG CTATTGGTGA AATGAAGCTA 480  
 TATGCCTGTT TACGTGGAAT GCTTTCAATT ACGAACAAAT CATTACTTAG TTTATTTAAT 540  
 15 GATAATGCTG AATTTAGAAA AAACGTAAAC TATCATTTAG CTTTGATTGG tAAAAC TGTT 600  
 aTCCAACATT TCCAGACCAA ATCACACCGG kAAACGTCCA AATATATTGn TCTATGCCCCG 660  
 GTTTACCACT ATC 673  
 20

## (2) INFORMATION FOR SEQ ID NO: 493:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1240 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 493:

AAAAAAAGTT AATGCTGACG GTGTATTAAC TTTTGATATT CTAGAAAATA AATATACTTA 60  
 TGAAGTTATT AACGCTATAG GGAAAAGATG GATTGTTAGT CATGTCGAAG GTGAAAACGA 120  
 35 CAAGAAAGAA TATGTAATAA CTGTCATTGA TAGGAAATCA GAAGGCGACA GACAACTGGT 180  
 TGAATGTACT GCTAGAGAGA TTCCCATAGA CAAGTTAATG ATTGATAGAA TTTATGTTAA 240  
 TGTAACAGGA TCTTTTACAG TAGAAAGATA TTTTAACATT GTGTTTCAAG GTACTGGAAT 300  
 40 GCTTTTIGAA GTCGAGGGCA AAGTTAAATC TTCAAAGTTT GAAAATGGTG GTGAAGGCGA 360  
 TACAAGGTTA GAAATGTTTA AAAAGGGATT AGAACATTTT GGTTTAGAAT ATAAAATAAC 420  
 GTATGACAAA AAGAAAGACA GATATAAGTT TGTATTGACG CCTTTTGCAA ATCAAAAAGC 480  
 45 GTCTTATTTT ATTTCTGACG AAtCAACGCC AACGCTATAA AACTCGAGGA AGATGCAAGT 540  
 GATTTCGCCA CCTTCATTAG AGGATATGGT AATTATTAG GAGAAaGAAAC ATTGCAACAC 600  
 GCTGGGCTCG TAATGGAAGC TAGAAGTGCA TTAGCTGAAA TATACGGCGA CATCCACGCA 660  
 50 GAACCATTTA AAGATGGTAA AGTGACTGAC CAAGAAACTA TGGATAAAGA ATTACAATCG 720

55

TATCCAGAAG CAGACCCACA ACCCGGAGAC ATAGTACAAA TAAAATCTAC CAAACTAGGT 840  
 TTGAATGATT TAGTCCGTAT AGTACAAGTT AAAACGATTA GGGGTATAAA CAATGTAATT 900  
 5 GTTAAGCAAG ATGTAACGCT TGGTGAGTTT AATCGAGAAC AACGATATAT GAAAAAAGTT 960  
 AATACTGCAG CTAECTATGT TTCTGGATTA AATGATGTTA ACCTTTCTAA TCCTAGTAAA 1020  
 GCGGCAGAAA ACTTGAAGTC TAAAGTAGCG TCAATAGCTA AATCAACACT CGATTTGATG 1080  
 10 AGTAGAACTG ATTTGATTGA AGATAAACAA CAGAAGGTAA GCTCTAAAAC TGTGACTACA 1140  
 TCTGACGGCA CTATCGTTCA TGATTTTATa GATaAATCma ACATTAAaGA TGTAAAaCG 1200  
 aTTGGAACGa TTGGCGATtC TGTAGCTAGA GGATCACATG 1240  
 15

(2) INFORMATION FOR SEQ ID NO: 494:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1311 base pairs  
 (B) TYPE: nucleic acid  
 20 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 494:

25 ACGGTGGATT TAGCCTAGAA GATTTAACGC ATCAAGGTAA ATTaTCAGCA TTTAGCTTTA 60  
 ATGATCAAaC AGGTCAAGCA ACATTGATTA CTAATGAAGA TGAAAACTTC GTAAAAGATG 120  
 30 AGCAACGTGC TGGCGTAGAT GCAAATTATT ACGCTAAACA AACATATGAT TATTACAAAG 180  
 ACACATTTGG TCGTGAATCA TATGaCAACC AAGGTAGTCC AATTGTTTCA TTAACGCATG 240  
 TTAATAACTA CGGTGGTCAA GATAACAGAA ATAATGCCGC ATGGATCGGT GACAAAATGA 300  
 35 TCTATGGTGA TGGTGATGGT CGCACATTCA CAAGTTTATC GGGTGCAAAT GACGTAGTAG 360  
 CACACGAATT aACACACGGT GTGACACAAG AGACAGCGAA CTTAGAATAT AAGGACCAGT 420  
 CAGGCGCTCT AAATGAAAGC TTTTCAGATG TTTTGGATA CTTTGTAGAT GACGAGGATT 480  
 40 TCTTAATGGG TGAAGATGTC TACACACCTG GAAAAGAGGG AGACGCTTTA CGCAGCATGT 540  
 CAAACCCAGA ACAATTTGGT CAACCAGCTC ATATGAAAGA CTATGTATTC ACTGAAAAG 600  
 ATAATGGTGG CGTACATACG AATTCTGGAA TTCCAAATAA AGCAGCTTAT AACGTGATTC 660  
 45 AAGCAATAGG GAAATCTAAA TCAGAACAAA TTTACTACCG AGCATTAAACG GAATACTTAA 720  
 CAAGTAATTC AAACCTCAAA GATTGTAAAG ATGCATTATA CCAAGCGGCT AAAGATTTAT 780  
 ATGACgAGCA AACAGCTGAA CAGGTGTATG AAGCATGGAA TGAAGTAGGC GTGGAGTAAA 840  
 50 AATATATAAA CaAGAAGAAG TAATGTAAA CACTTATAAA TAATTAAATT TTAAATACAG 900

55

ATTAGATGAG AGGAGTGTGA GGGTTGTCTG CCGAAAGACT ACTCGGCAGT CTAAAATCAT 1020  
TACAAGTAGT AGATATGTGA TAATTAAATG CTGACTTAGA ATACAAAATT CATTTTAAAA 1080  
5 GTTGTCAACAA AAAATTTACA TGTATTTTTA TTATCTTTTG CAAAACAAAG TGTAAATTA 1140  
TAAATGAaAC ATGCATGAAT TTATTTTTTA ATACAAGAAA CGTAACTACC AAAGGAGTTT 1200  
ACAATATGAA GAAAAGTAAA CGATTAGAAA TTGTTTCTAC AATAGTTAAA AAGCATAAGA 1260  
10 TTTATAAAAA AGAACAAATc ATTTCATATA TTGAAGAATA TTTTGGTGTA A 1311

(2) INFORMATION FOR SEQ ID NO: 495:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 1761 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 495:

TGCACTTTCT AAAAATACTT GCTTTACTTG TTCCAATTC TTGTCAGACA TTGATTCGGC 60  
25 TAATTTATTC ATTATTGCCT CAAATACGTT AATTATGTCA TCCATTTTCAG TACTATAAAA 120  
ATCAAACCAT TTGCCGTAT CTTTTTCTCT ATTAAGCTTA TGATCACTTT GAGAGCGCTT 180  
AGCTAATTCT GCATAAATAT ATGGACAAGG TGCCATTGCA GCAATTGTAT AAATAGCATT 240  
30 TTCACGACTA TGCGCTTGGA AATACATATG TTTTATGTAA TGGTCGCCAC TTGGAGGCCA 300  
AACTTTTGTT TTAATGATTT CTTCGTATGA TTCACCAACA ATTTGCGCTA AAATATCATG 360  
CGCAAGTACT TCACCTTCAA CCATAAATTC TATTTGCTCT ACTAAAAATT TTACGTCATT 420  
35 CATGCTATTC ATTTTGGGAA TTAACAAAGC ATATAAGTTT GTAAATTCTT TTAAATACGC 480  
AGCATCAGCT TTAAAGTAAT GGCCTAATGC GTCAGcTCCT ATATCTCCGG ATAACATCTT 540  
CTGAATAAAG TCATCCTCAT AAATATCATT AATGATTGGC TTTGCAGCTT GGTACAATTT 600  
40 TTGTGAAAAT TCCATTGTAA AAAATCCTCC CTAAATAAAA AACTACTTTC CAACATGAAA 660  
GTAGTTTGAT GGCAATGTTG CTATACTAGC CCCATCACTT CAATAACTAC TTTCTACGT 720  
TGGTACTAAC CAAATCAGGT CATAAGGGTC TGAACAATTC ATCTCAGCCA TATCATTAGG 780  
45 CTCCCCTAGT AGTTCCTTAG TATTCAATTG CAAATTAATC TTAGCAAACG GTTCAACAT 840  
TTTCAATTAT TGTGCTCAG TTGTATTATT ATCTTTAAAT AATAATTCTA TAATGACATA 900  
TATTTGCGAA ATAAAAAAC CGGAACATAT CGAGAATTCC CCGATATATT CCAATCTAAA 960  
50 AGTTACTTAT ATAACATTA ATTAGCTATG CATAAATGGC TTATGCAGTA ACCCAATGTC 1020

55

TTGCTGGTGA TACACCTTTA TATTTAGCAG GTGCTACTGA ATCCCAAGTT GATTGTAAGA 1140  
 ATTGATACCT ACCAGCTGCA CCTGgATGTT GGrTTTACAG CATGAATATT GCCACCTGaT 1200  
 5 TCACGTTGrG CAATTTGTTT TAGATGAGCa TTCACATTTA CTGATGAACC TTCTGATGAT 1260  
 TTTGATyCAG TTGGTGTTC AGTAACTTGT GAATTGTTTG ATGTTGATGC TTGTGGTTGT 1320  
 TGAGTTTGAG CATTTTGTGG TGCTTCAACT TCTTGTGATT GTACTTGATT AGCTTGAACA 1380  
 10 GCTGATGGTG CAACATTATT AGTTCAGGT GCTTGTGCAC TCATGTCTGC TCCATTAGTA 1440  
 CCTGTTGCAT GGTAATTCCA AGCAAAGTGT GTACCATCTG ATTCAAAGTG ATAAGTAAAC 1500  
 CCTTCATAGT CAAATGTATA ATTATAAGCC CCAGCTTCAA TTGGTTTTTG ATTTAATGTT 1560  
 15 TGATCATTTG ATTGCGCCAT TTGCCTGAAA GATGCTTTAT TTAAGTCCGC TTCACnTGCA 1620  
 TGGGCTTCGT GGACCTGCAT TTCCTGGCTA CGATTCTTAA ACCTACTGGC nAAnGATGAT 1680  
 GCGAGTAATG TTTTCTTCAT AATCTTAAAA TCCTCCTACA AGTGAATTTG TGTCTCTAAA 1740  
 20 AGTTTTACAG TGGACGACTG T 1761

## (2) INFORMATION FOR SEQ ID NO: 496:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 794 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 496:

TCATTTATGA AAAATGTCCG AnAGAnCCaa GaAAmaCAAT TGAgCGTGAA GAAAAAGCAA 60  
 35 GACTTAAAGA AGAACAAAAG GCACGTCAAA ATGAACAGCC ACAAATAAAA GATGTGAGTG 120  
 ATTTTACGGA AGTGCCTCAA GAAAGAGATA TTCCAATTTA TGGGCATACT GAAAATGAAA 180  
 GTAAAAGCCA GAGTCAACCA AGTCGAAAAA AACGAGTGTG TGATGCAGAG AATAGTTCGA 240  
 40 ATAACATCGT AAATCATCAT CAAGCAGATC AGCAAGAACA ATTAACAGAA CAAACTCATA 300  
 ACAGTGTTGA AAGTGAAAAC ACTATTGAAG AAGCTGGTGA AGTTACGAAT GTATCGTATG 360  
 TTGTTCCACC GTTAACTTTA CTTAATCAAC CTGCAAAACA AAAAGCAACA TCTAAAGCTG 420  
 45 AAGTGCAACG TAAAGGACAA GTACTAGAGA ATACATTAAA AGATTTTGGG GTAAATGCAA 480  
 AAGTGACACA AATTAAAATT GGTCTGCAG TAACTCAATA TGAAATTCAA CCAGCTCAAG 540  
 GGGTTAAAGT GAGTAAAATT GTAACTTGC ATAATGATAT TGCATTAGCT TTAGCAGCAA 600  
 50 AAGATGTTAG AATCGAAGCG CCAATACCTG GTCGTTCTGC AGTAGGTATT GAAGTGCCAA 660

55

ATAAACTAGA AGTTGGATTA GGaAGAGATA TATCAGGTGA TCCAATTACT GTTCCACTAA 780  
 ATGAAATGCC ACAC 794

5 (2) INFORMATION FOR SEQ ID NO: 497:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1161 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 497:

15 AGCCAGTTTT GcATTTCGTC AAATCGCAAT GAATATATTG ATTGCTTGTC AAAATTTGGA 60  
 AGAAATTTAT TCTTCTTTTT CTTTTCAGCA GTTATCATTC CTAATTGTCC TTTCTTTTTTA 120  
 TCTTAGTGAT AAAGAATCCA TCTGAATTAA AGTCTTGCGG CATGATTTGT AACGTTTGA 180  
 20 CCAACTCTCC AGTTATCGGA TGTGAAACG GTTCAAATTC GAAGTTTTTA TTATTTTCA 240  
 AAAACGTATA AATCACGTTT TCATTTTCTA GTTGCTCAAT TGTACATGTT GAATAGATGA 300  
 TTTCTCCACC TATTTTACA TTGTTTTTA CATTTTCAA TATTCAAGC TGTAATTCAA 360  
 25 CTAGTGACTC AATATGTTGT TTGCTTTGAG TATACTTAAT CTCCGGCTTA TGTCTCATT 420  
 CACCTAATCC GCTACATGGT GCATCAACAA GTATCTTATC GTATGTTTTA TCATAAGGTT 480  
 TTGTCGCATC ATGTTGAAAA GCTTTAATAT TTGTTAATCG TAATTTTAT ATATTAAAA 540  
 TAATTAAGTC TATTTTGTGA TCATGTATAT CTGAAGCGTC AACTTGCCCT TCTGGCATT 600  
 AACTTCAGC AATGTGACAA GCTTTACCGC CAGGTGCACT ACATGCATCT AATACGTGAT 660  
 35 CATGTCGGTC TACATTCATA ATGTTGCAa CAAACATTGA GCTTTTATCT TGAATTGAAA 720  
 CGAATCCATC TTAAATGAA CGAGAATGAA TAATTGGTTG TCCTCCTATA TGGAGACAAT 780  
 AAGGTAAGTC ATGATCTTTT TCAACGTCAT AACCTTCGTC TTGCAACTTT TCAATAATAT 840  
 40 CATCTAATGA TGCTCGCGTC AGGTTGGCAC GCACAGTTGT TGATGTCGTT TCTAAAAATG 900  
 ACTGTAAAAT TTTTTCAGTT TCTTCGAGAC CATAATGTGT TGCCCAATGA TCTATAATCC 960  
 ACTTCGGCAT ACTATACTCG ATTGCCATTC TTTTTTIAGG ATCTGCAATT TCATTAAAA 1020  
 45 CAGGTAAGTC ACTACGCATC ATTGTACGTA AAATACCATT TACGACATTA CCATTATGAT 1080  
 AGCCACCGCG TTCTTTTGCT ATTTCAACTG CTTCAATTAAT AATGGCATGA TTTGGAAC 1140  
 TATCTAnATA nACATATTGA T 1161

50 (2) INFORMATION FOR SEQ ID NO: 498:

55

(A) LENGTH: 1504 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 498:

10	AGCTCACGTC ATCTTCGGCG GCGCTAAATT AAAATAATCA ATTTCTGAGT TAAACTTTTA	60
	TTTACAACAT ACTATTACTA TACATTACAA ATTTTAAATA TATACATTAC ACTCATTACT	120
	CAATGGAAAG CGTATGATTT CmcAGCCCCC CTAGCTTGTA GAAATCATAC TTTCCTTTTT	180
15	TCAATATATA TACAACATTT AAATCCCATG AGATTGCAGA GCACATAAGT AAATTTTTTT	240
	AGAGCTTGAG GTTTGTTTAG CTTAAGCAAC CCATGAGCTC AAACACTTCC TGTTACACATA	300
	ACACTACAAA TCGCATTATG TTGCTTAATC TTATGTTTAT ATAAATTACA CACAATAAAT	360
20	AGAAAGAATG TGAACATCAT GAATAAATTA TTGCTACTCG TTACATTATC CATTCTGTGT	420
	GGTTCAGGTA TTGTTATGTT AATGCAAGGC TACGAAAAAT TAACGGGCGG ATTTACGCTG	480
	AAAGGTTTAG TACCAATCAT CGCTAACAAAT ACTGATTCAC CAGAGTGGTA TAAGTGGTTT	540
25	TTCGCAAATA TAGTTGCACA TACGACGTCA TTATTTGATA TTGTTGTCCC ACTCGGAGAG	600
	ATTGCAATTG GATTAGGTTT AATTTTTGGA GTTTTTGCAT ATGCTGCTAG TTTCTTTGGA	660
30	GCCTTTGTTA TGATAAATTA TATCTTAGCA GATATGATAT TTACGTATCC TCTTCAATTA	720
	ACTTTCTTTA TCCTTTTACT AATGAGTCAC TCATTGTTAA AACAGATTTC ACTTAAAGAA	780
	ATCATTAAAT ACTTTAGAGG TCGTAAGAAC AGAGGTGAAA AAATAGATGA CCCACTTACT	840
35	GATCGTGGAT GATGAACAAG ACATTGTAGA CATTGTGCAA ACCTATTTTG AATATGAAGG	900
	TTACAAAGTA ACAACGACAA CTAGCGGTAA AGAAGCAATT TCTTTACTAT CAAATGATAT	960
	TGATATCATG GTACTTGATA TCATGATGCC AGAAGTTAAT GGTTACGACA TTGTCAAAGA	1020
40	AATGAAAAGG CAAAATTAG ATATCCCCTT TATCTATTTA ACTGCCAAAA CACAAGAACA	1080
	TGATACCATT TACGCCTTAA CTTTAGGTGC AGATGACTAT GTCAAAAAAC CATTTAGTCC	1140
	AAGGGAACCT GTTTTACGTA TTAATAATTT ACTTACAAGA ATGAAGAAAT ACCATCATCA	1200
45	ACCAGTTGAA CAACTGTCGT TTGATGAATT AACACTTATT AACTTAAGTA AAGTtGTGaC	1260
	tGTAAaTGGT CACGAaGTCC CTATGCGTAT TAAGGAATTT GAGTTATTGT GGTATTTAGC	1320
50	TTCTAGAGAA AATGAAGTTA TTtCTAAATC AGAATTACTT GAAAAAGTTT GGGGATATGA	1380
	CTATTACGAA GATGCTAATA CCGTGAATGT CCATATACAC CGTATTAGAG AnAAATTAGA	1440
	AAAAGAGAGC TTTACAACAT ATACCATCAC AACTGTATGG GGATTAGGAT ATAAATnTGA	1500

55

## (2) INFORMATION FOR SEQ ID NO: 499:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1623 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 499:

ATTGAAAGCG ATAATTCGTA nTAATTGAGT TTGTTGAAAA ATTTAGGGTA ATGTAAAGAT 60  
 ATAAAAGATA CATAGAyTGG AGAGATATAA AGATGTTGAA TGAGATACAA ATATTAAaATA 120  
 aTGGATACCC GATGCCCTTCA GTTGGGTTAG GTGTTTATAA AATCTCTGAC GAAGATATGA 180  
 CTAAAGTTGT AAATGCTGCA ATTGACGCAG GCTATAGAGC GTTTGATACA GCATACTTTT 240  
 ATGATAATGA GGCTTCACTA GGACGAGCAT TAAAGGATAA TGGCGTCGAT AGAGAAGATT 300  
 TGTTTATAAC AACGAAGTTA TGGAATGACT ATCAAGGTTA TGAGAAAACA TTCGAATATT 360  
 TCAACAAATC GATTGAAAAT TTACAAACTG ATTATCTTGA TTTATTTCTA ATACATTGGC 420  
 CTTGTGAAGC AGATGGTCTA TTTTtagaaa CATATAAGC TATGGAAGAA CTTTACGAGC 480  
 AAGGTAAGGT AAAAGCAATA GGTGTATGTA ATTTAATGT TCATCATCTA GAAAAATTAA 540  
 TGGCTCAATC AAGTATCAAA CCAATGGTGA ATCAAATTGA GGTACATCCA TATTTTAACC 600  
 AACAAGAATT ACAAGAATTT TGTGATCGTC ACGATATTAA AGTGA CTGCA TGGATGCCTT 660  
 TGATGAGAAA TAGAGGACTA CTAGACGACC CTGTCATTGT TAAAATTGCT GAAAAATATC 720  
 ATAAAACACC AGCACAAGTT GTATTACGTT GGCATTTAGC ACACAATAGA ATTATTATTC 780  
 CAAAATCTCA GACACCTAAA CGCATTCAAG AAAATATAGA TATTTTAGAT TTTAATTTAG 840  
 AATTAACAGA AGTAGCTGAA ATTGATGCTT TAAATAGAAA TGCAAGACAA GGTA AAAATC 900  
 CAGATGATGT GAAAATTGGG GATT TAAAAAT AACTGGATGT TAAATTTTAC GTTTATGAAT 960  
 GCCTTTTAAT GTGTACATTA AAATAAATGA GTTGGTTTTT ACTATTTGAT AAAACAATAC 1020  
 TCAGGTACAT TCAAAATCTT TTAAATAAAA AGGATGGACA TAGATGAAAA TTAGAGTCGT 1080  
 CATTCTTGT TTTAATGAAG GGGAAGTCAT TACACAAACA CATCAACAAT TAACTGAAAT 1140  
 ACTTTCACAA GATAGTAGTG TGAAAGGCTA TGATTATAAT ATGCTTTTCA TAGATGATGG 1200  
 TAGTACGGAT ACCACTATAG ATGAAATGCA ACATCTGCC ACAATAGATA GGCATGTCAG 1260  
 CTTTATTTCT TTTAGTAGAA ATTTTGAAAA AGAAGCAGCT ATGATTGCAG GTTACCAGCA 1320  
 TAGTACTGAA TTTGATGCAG TCATCATGAT AGATTGTGAT TTGCAACATC CACCTGAATA 1380



TAGAAGTGGT GAAAATTTTA GTCGCAAAAC ATTAAGCCAT TTGTATTATA AGTTAGTTAA 1500  
 TTGCTTTGTA GAAGAAGTAC AATTTGATGA TGGTGTGGT GATTTTAGAC TTTAAGCCA 1560  
 5 AAGAGCTGTT AAATCCATTG CATCACTGA AGAATATAAT CGnTTTCAA AAnGGnTATT 1620  
 TGA 1623

## (2) INFORMATION FOR SEQ ID NO: 500:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 605 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 500:

20 AAAGTnGGTG AAnCTATATA CTTAATCTAT ATTTATATAT TAACCATTAG GGTAAAAAAT 60  
 TACTCTAGCA TTTATGAATA GATGGGAGTT TATTTTATTA TTATATAGGA GAGATGTTGA 120  
 ATGACACATC GCGCACTATT AGTTGTTGAC TATTCATATG ACTTTATCGC AGACGACGGC 180  
 25 TTACTAACAT GCGGTAAACC TGGACAAAAT ATTGAAGATT TTATTGTTTC TCGTATCAAT 240  
 GACTTTAATT ATTATCAAGA CCATATATTC TTTTTrrTGG ATTTACATTA TTTACATGAC 300  
 ATTCATCATC CTGAAAGTAA ATTATTCCCA CCACACAATA TCGTAGATAC AAGTGGTAGA 360  
 30 GAATTATACG GTAAAGTAGG TAAATTATAC GAAACAATTA AAGCGCAACC TAATGTACAT 420  
 TTCATTGATA AAACGCGCTA TGATTCGTTT TTTGGTACCC CGCTTGATAG TTTATTGAGa 480  
 GAAAGAAGTA TTAATCAAGT CGAAATCGTT GGTGTATGTA CCGATATTG CGTGTTACAT 540  
 35 ACAGCAATTT CTGCATACAA CTTAGGtTAT AAAATTTT CAG TACCTGCTGA GGGAGTGGCT 600  
 CATT 605

## (2) INFORMATION FOR SEQ ID NO: 501:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1739 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 501:

50 TAGGTTnAAA GCATAGnTTT nTCAAAAAGA CAAATCATT C ATATATTGGA GGATATTTTG 60  
 GTGTAAGATA TAGTGCAACC ACAATTGCTA AAGACTTGAA GGAACATAAT ATATATCGTG 120

	5	10	15	20	25	30	35	40	45
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110
111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130
131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150
151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170
171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190
191	192	193	194	195	196	197	198	199	200
201	202	203	204	205	206	207	208	209	210
211	212	213	214	215	216	217	218	219	220
221	222	223	224	225	226	227	228	229	230
231	232	233	234	235	236	237	238	239	240
241	242	243	244	245	246	247	248	249	250
251	252	253	254	255	256	257	258	259	260
261	262	263	264	265	266	267	268	269	270
271	272	273	274	275	276	277	278	279	280
281	282	283	284	285	286	287	288	289	290
291	292	293	294	295	296	297	298	299	300
301	302	303	304	305	306	307	308	309	310
311	312	313	314	315	316	317	318	319	320
321	322	323	324	325	326	327	328	329	330
331	332	333	334	335	336	337	338	339	340
341	342	343	344	345	346	347	348	349	350
351	352	353	354	355	356	357	358	359	360
361	362	363	364	365	366	367	368	369	370
371	372	373	374	375	376	377	378	379	380
381	382	383	384	385	386	387	388	389	390
391	392	393	394	395	396	397	398	399	400
401	402	403	404	405	406	407	408	409	410
411	412	413	414	415	416	417	418	419	420
421	422	423	424						

(2) INFORMATION FOR SEQ ID NO: 502:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1745 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 502:

5	CTGTACATAC AGCAATATCG TTAACAAATG AAAACAGTAT TTTAGGATTG TAAACCATGA	60
	TAAACCTAAA ATACTGTTAT TTTTATTACT TAAATTTCTT CTTCAATGCC TTTTCAACAT	120
	AAGGTGGAAC GAATTCAGAA ATATCTGCTC GATAAGCTGC AACTTCTTTA ACAATACTTG	180
10	AACTTATAAA TGAATAATTA GTACTAGACA TCATATATAA CGTTTCAATT TCATTGTTCA	240
	ACTTTTATT CATTGAAGTT AAGCGTAATT CATATTCAAA ATCACTGACT GCTCTTAAAC	300
15	CACGTATGAT TGTTTTAGCT CCTACTTGTT CACAATAATC GACTAGTAAA CCACTAAATT	360
	GATGAACCTT GACATTAGGT AAATGTTTAA CAGATTGTTC AATTAAATCC ATACGCTCTT	420
	CTAAACTAAA CGTACCTTCT TTTTACTAT TTTAAGAAC ACAGACATGA ATTTCATCAA	480
20	ATCTATCTGT ACTTCTCTCA ATAATGTCTA AATGACCATA AGTAATGGGG TCAAACTAC	540
	CCGGAATGAC CGCTATTGTA TGTTCATGC TATTCTCCCT TTTCTAATAA CAATGTGTCT	600
	GTCAACCCAT AATGGTAACG TTTAATCATA TTAAACGGT GATAATCTAT TTCTTCATGA	660
25	TTGCTAAATT CACAAACGAT GATACCATT TCTTTCAATA AATTAACTC TGAAATTAGT	720
	TTTAAAGCTT TATCAATGAG ACCTTTATTA TAAGGTGGAT CTAaGAAAAT GACATCAAAT	780
	TGAATATCAC GTTTTGACAA TGCTTTTAAA GCTCTATCTG CATTATTTT ATAACTTCA	840
30	GATTGTGCCT CTAAATCCAA ATTGCAAGA TTGATTTTAA TAACTTTTAC AGCTTTAAAA	900
	TTTTGaTCAA CAAAGATTAC CTTATCCATA CCTCGAGAGA GTGcTTCTAT TCCAAGCGCC	960
35	CCGCTTCCTG CAAATAAATC TAAACCTATA CCTGACACAT CATATAAACT ATTAAAGATA	1020
	CCTTCTTTAA CTTTATCCAT AGTTGGTCTC GTATTACGGC CTTCCATACT TTCTAAAGCT	1080
	TTACTTTTAT GTTTACCTGC AATGACGCGC ATGTTGTTCA CACTTCCAAT TCATTTAGTT	1140
40	ATTTAATATA ATTTATTGAG AAAAAGGAGA ATGATAAACC AATGAAACAA ACATTTATTA	1200
	CACTTGGTGA AGGTCTAACA GATTGTTCG AATTCATGAC GATGATTGAA TATAACCATC	1260
	AACGTATTGA TAAAATTATC TATTTTCATT CACCACAAGC TGAAAATAAA AAGTCATCTG	1320
45	TAGCAATCAT TATGAACCCT ACAACTGGCA ATCATTTCCA AGCATTTTAT ATCATGATAA	1380
	ACGCTATTAA ATATCCATAT CCAGATTCAA ATAAAAAGTT TCAAATGATA AATGATTGTG	1440
	CTGAAAAATT CGACATACCA ATTTTAGGTA TCGATGTACA GCCCCCTCAA GCATTTCATG	1500
50	ATTTATCGTT ATATTATAAT TATTTAATTA GTGTGTTAAG GCTCCAAAA TGGATACCAG	1560
	aACTTCAATA ATATTAATTA TATATTCGT GTTCTCTTT TTCGTAAGTT TTCTTTAAGT	1620

55

TTTGATCAAC ATCTTGCTCA TTCACATACA TAATTACAAA TTTACGATCT CTATTTGAAT 1740

GAACG 1745

(2) INFORMATION FOR SEQ ID NO: 503:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1035 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 503:

TCGTCTTTAA TCTTGCTGAC TTTAGAAGGC TTACGAGTGC GGTTACCATT TTTTGCATCT 60

TTAACTGATT GAACTAAAGC TtGACGTGTA GATTTATCAG CTAAACTAAT TGCACCACCA 120

ATTACGGCAC CAATTAAAAT ACCAGGAACA AATTATTTTT CCATAAAAAA CTACCCCTCT 180

TTCAAATTG CATCTTTTAC GATGTAGTCT ATTAAATTAT CACAAGATGA TAATACCATG 240

TCGTATACAC CTTCAAAATT ATTCGTGTAG TATGGATCTG GTACATCACT CTCTTCATA 300

TTACTAAATT CTAACAGTTT GAACAATTGT CCCTTAAGAT TAGGATTGAT AGATTTAATA 360

TTATCAACGT TACTTTTGATC CATAGCCACA ATGTAATCAA AATCATCTGT CGCTTCGAAT 420

AATTCATAA TCATGCCATC AAATGGAATA TTGTGTTTGT TGAGAATTTT TTGTGTACCT 480

TCATGAGGTG GCTCTCCTAA ATTCCAGCTA CCAGTACCTC TTGAATGTAC TTAAATATCA 540

TGAATATTTC TGTCTTTAAG TCTTTGTGCG ATGATTGCTT CTGCCATTGG AGAACGACAT 600

ATATTGCCAA GACAGACAAA TGCTACATCT ACCATTTTGA TTCCTCCAAA CTATGTAGTT 660

ATATCCCAT TTTATAGCGA CTTTAAACAA TAAGAAAGCA GATTATATAA AATTCTATTA 720

AAGTTTATTA AATTGTGATA CTTTGATAAC ATA ACTATTA TTAGAGGTGA ACATTGTGGC 780

TATGACAAAT GAAGAGAAAG TnTTAGCTAT TAGAGAGAAG TTAAATATtG TTAATCAAGG 840

ATTATTAGAT CCTGAAAAAT ATAAAAATGC AAATGaAGAA GAATTAACAG ATATATATGA 900

TTTTGkTCaA yCAAGAGaAA GATTGTGCGC AAGTGaAGTG mCAGCTATTG CTGaCGCTTT 960

AGGACAATTG CGACACGAAT AGGAGTGGGA ATTTTGACTA ATTACAAAGA AAAGTTACAA 1020

CAATACGCTG AACTA 1035

(2) INFORMATION FOR SEQ ID NO: 504:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1284 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 504:

5 AnCCTGACAA GATTCTTATA TTCATAGATA TGAGAGCTAA AAATGAAATC AACAAACaNa 60  
 CAATAAAGTA AACGATAATA GCCCATATAC CATTTTGTAA CCATATTACA AATTGTGTTG 120  
 TATTATAGCC ATTTCCAGCT AATAATTGCT GGATAAATGC ATTATTGTTT AATGTATTTT 180  
 10 CAAGATTAGC AAtCGATGTG TTATTACTGA ATGAAACAAG TGCTATAAAC ATCGTAATGA 240  
 CAGTAAGCAC TAATAACATC ACCCAACATA ACCAACCTAG AACTTTTTCA GTTAATCTAC 300  
 TTACTGGACG TTTAATTIGA GTAAATTGTT CTCCAGTCAT TCGTTACAAC TCCTTATAGT 360  
 15 ACTTATCCCG TTATTATAAC TAAATATACA GTAAATAACT ACTATTTATG ATTTTATTTT 420  
 AATGACATTT TGAAATTCAA AAAGTTTTCA TTGTATTCAC TTAAAACTTC AGGTCCTAAA 480  
 20 TCTTTATAAA CTTCAAGGCG TTCTTGCTCT TTCTTAGTCG GATAAAAACG ATGGTCGTCT 540  
 TTAATCTCTT TAGGCAACAA TTGTCGAGCA GCCTTGTTTG GCGTTGCATA GCCTACGAAT 600  
 TCTGTATTTT GCTTGTTATT TTTAGCATCT AATAAAAAAT TCATAAATTT ATATGCACCC 660  
 25 TCTTTATTTT GTGCCGTTTT TGGAATTACC ATATTGTCGA ACCATAAATT CGATCCTTCT 720  
 TTAGGAATAA CATAATTATA TTTATCCCCT TCTTGCACTA GAGGTGCTGC AACACCACTC 780  
 CAAACAACCG CTATGTTACC TTCATTTTGT TGAAGCATCA TGGTAATTTT ATCACCTACG 840  
 30 ACACCTCTTA CTTGTGGTGC TAGTTTGGTT AAATCTCGCT CTGCTTCTTT TAAATGGTGC 900  
 GAATTACGGT CATTAAGATT ATACCCAAGT TTATTCAAAC TCATGCCTAT AATCTCTCTA 960  
 GCACCGTCAA CTAGTAAAT TTGGTTTTTA AATTTAGGAT TATACAATGA CTTCCAACCTA 1020  
 35 TCAAATGATT CATTTGGATA CTTTCTTTTA TTATATAAAA TACCTACAGT TCCAAAGAAA 1080  
 TAAGGTAAAG AATATTTATT GCCTCTATCA AATGACATAT TCATATAATC TGAATCTAAA 1140  
 40 TTTTAAATAT TAGGTACCTT ATTATGATCT ATTGGTAACA ATAAATGATC TCTTTTCAAT 1200  
 TTTTGAAGTG nATATTCACT AGGAnAAGCA ACATCATAAT GTGTACCGCC AGTGCGAATT 1260  
 TTGGnGTCCA nCGCTTCATT TGAA 1284

## (2) INFORMATION FOR SEQ ID NO: 505:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5763 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

	ATAATTAACA TCTTTTGCTA TATACCACCA GTTTGATACA TAAAATATCG CAGCAATAAT	60
	ATCATGTTTA ACCCTAATGA TATTATCTGA TTTTAATAAT AAGGTTGCTG TCCCTACAAC	120
5	CATTAATAAA ACTATGACTG CTGGTAATAA ACGTTTTAAA CGACGTATCC AAAAGCTTTT	180
	CAATTTAATG ATACCTGTGT CATCATACTC TTTGAGTAAT AAGCTTGTA TTAATAAACC	240
10	AGAGATCACA AAAATGTAT CCACACCTAA AAAGCCACCT GtCAACCATT GCTTATTTAA	300
	GTGGTAAATA ATAATTCCTA GAACAGCGaT TGCCCTCAAA CCATCGAGCC CTGGAATATA	360
	TCTCATTTTC TTATACTTTG TAAAACCCCTT TGTTTTGTTT ATTTTTCAT TCTTCCCTTT	420
15	TAAAACGTGT CTTCTTAGAT GCTTAATTAA ATTTAGTTAT GCTGTTTAAA AGAATATTGA	480
	AATGCATATG TATATTATTG AATTACGACA TCATCAAAAT CATATTGACT AAAATACTGT	540
	TAAATTAAAA AAATTACCAA TGATGATTCT TACTTCGAAA TCCAATTTGT AATGCAACTC	600
20	GGCAACTTAA AAACATGAA GTATTATGTA TTGTAATATA ACTGTAATAT AAATTCAATT	660
	TATTATAAAA ATTTTCAAGA AAATATTCAA CTAGAAAATG AATTGTGCAC TCTTGAAGT	720
	GCAAGTCACT GTCTTAATTC ATATTTTTTG AAACAAGTTA GATATAAATT TTCAAAATAA	780
25	AATCAGAAAC TAGAACATAA ATAAGGCTCC CTTCAAAATT TTCATTTTTC AATGTCTACT	840
	TTGAAGGGAG CTTATTCACA ATGAATTATA CTCTACAATG TTATATTGAC TGCGGGCCCA	900
30	AACACAGAGA ATTTGAAAA GAAATTCTAC AGGCAATGCA AGTTTATGTT AGCTCACACC	960
	AAGTGCAATC TTAGCGTAAC GTGACATCAT ATCTTTTGTC CAAGGTGGAC TCCATACGAT	1020
	ATTCACCTCA GTATCCTGAA TTTCAGGAAT CTCTGCTAAT ACTGTTTTAA CTTGaTCAAT	1080
35	AATTTGAGGT CCCATTGGAC ATCCATTGA TGTTAAAGTC ATATCAACTG TACATACGCC	1140
	TTCATCATCA ACATTCACCT TGTATACTAA ACCCAAATTA ACGATATCAA TTCCTAATTC	1200
	AGGGTCAATT ACCATTCTA ATGCACCTAA GATACTATCT TTCAATGCCT CTTCCATCCA	1260
40	TATCACCTCT TTAATGTCAT ATTATTCATA ATATATCAAA TATCCGACAA AACGCCAATA	1320
	AAATGCTATG ATGTATCTAT ATGAACTAAG CAACTTATGA GGAGAGAGAT ATGCAACCAC	1380
	ATTTAATATG TCTAGACTTA GACGGAACAT TATTAAACGA TAACAAAGAA ATTTTCATCAT	1440
45	ATACTAAACA AGTATTAAAT GAATTACAAC AACGTGGaCA CCAAATTATG ATTGCGACTG	1500
	GCAGACCTTA TCGTGCAAGT CAAATGTATT ATCATGAATT AAATTTAACG ACACCAATTG	1560
50	TTAATTTTAA TGGCGCTTAC GTACATCACC CTAAAGATAA AAACCTTCAA ACTTGCCATG	1620
	AAATTTTAGA TTTAGGCATC GCACAAAACA TTATTCAAGG ATTACAACAA TATCAAGTAT	1680
	CGAATATTAT AGCAGAAGTG AAAGATTATG TTTTCATTAA CAATCATGAT CCAAGATTAT	1740
55		

	AAGAATCCCC TACCTCAATT TTAATTGAAG CCGAAGAAAG TAAAATACCT GAAATCAAAA	1860
	ATATGCTTAC TCATTTTTAT GCCGATCATA TTGAGCATCG ACGCTGGGGC GCACCATTCC	1920
5	CTGTCATTGA AATTGTAAAA CTTGGTATTA ATAAAGCAAG AGGCATTGAG CAAGTTAGAC	1980
	AATTTTAAA TATTGACCGA AATAATATTA TTGCATTCCG TGATGAAGAT AATGATATTG	2040
10	AAATGATTGA GTACGCGCGT CACGGTGTTG CTATGGAAAA TGGTTTGCAA GAACTTAAAG	2100
	ATGTAGCGAA CAATATTACA TTCAACAATA ATGAAGATGG CATTGGTCGA TATTTGAATG	2160
	ATTTCTTTAA TTAAATATT AGATATTACT GTTAATTTAT AACTAATCAT TTTATAATAT	2220
15	TTTAAACAA TAGGAGGTAA GTTACGATGC CCAAATAGT CGTAGTCGGA GCAGTCGCTG	2280
	GCGGTGCAAC ATGTGCCAGC CAAATTCGAC GTTTAGATAA AGAAAGTGAC ATTATTATTT	2340
	TTGAAAAAGA TCGTGATATG AGCTTTGCTA ATTGTGCATT GCCTTATGTC ATTGGCGAaG	2400
20	TTGTTGAAGA TAGAAGATAT GCTTTAGCGT ATACACCTGA AAAATTTTAT GATAGAAAGC	2460
	AAATTACAGT AAAAATTAT CATGAAGTTA TTGCAATCAA TGATGAAAGA CAAACTGTAT	2520
	CTGTATTAAA TAGAAAGACA AACGAACAAT TTGAAGAATC TTACGATAAA CTCATTTTAA	2580
25	GCCCTGGTGC AAGTGCAAAT AGCCTTGGCT TTGAAAGTGA TATTACATTT ACACTTAGAA	2640
	ATTTAGAAGA CACTGATGCT ATCGATCAAT TCATCAAAGC AAATCAAGTT GATAAAGTAT	2700
30	TGGTTGTAGG TGCAGGTTAT GTTTCATTAG AAGTTCCTGA AAATCTTTAT GAACGTGGTT	2760
	TACACCCTAC TTTAATTCAT CGATCTGATA AGATAAATAA ATTAATGGAT GCCGACATGA	2820
	ATCAACCTAT ACTTGATGAA TTAGATAAGC GGGAGATTCC ATACCGTTTA AATGAGGAAA	2880
35	TTAATGCTAT CAATGGAAAT GAAATTACAT TTAAATCAGG AAAAGTTGAA CATTACGATA	2940
	TGATTATTGA AGGTGTCGGT ACTCACCCCA ATTCAAAATT TATCGAAAGT TCAAATATCA	3000
	AACTTGATCG AAAAGGTTTC ATACCGGTAA ACGATAAATT TGAAACAAAT GTTCCAAACA	3060
40	TTTATGCAAT AGGCGATATT GCAACATCAC ATTATCGACA TGTCGATCTA CCGGCTAGTG	3120
	TTCCCTTAGC TTGGGGCGCT CACCGTGCAG CAAGTATTGT TGCCGAACAA ATTGCTGGAA	3180
	ATGACACTAT TGAATTCAAA GGCTTCTTAG GCAACAATAT TGTGAAGTTC TTTGATTATA	3240
45	CATTTGCGAG TGTCGGCGTT AAACCAAACG AACTAAAGCA ATTTGACTAT AAAATGGTAG	3300
	AAGTCACTCA AGGTGCACAC GCGAATTATT ACCCAGGAAA TTCCCTTTTA CACTTAAGAG	3360
50	TATATTATGA CACTTCAAAC CGTCAGATTT TAAGAGCAGC TGCAGTAGGA AAAGAAGGTG	3420
	CAGATAAACG TATTGATGTA CTATCGATGG CAATGATGAA CCAGCTAACT GTAGATGAGT	3480
55	TAAGTGAGTT TGAAGTGGCT TATGCACCAC CATATAGCCA CCCTAAAGAT TTAATCAATA	3540

	GTTAGAATTA TGTTGGACTG GTACTACTAT CCAGTCCATT TTTTATGTTT AACATTTTTTA	3660
	GAATCAAAAA AGACATAAGG TCTTGGACTA ATAATTGTCC ATGCCTTATG TCATATACTA	3720
5	TATGTCTTAT CAATTAGCCA ATACCGAATA ATTTTGATAT AGGsCCTAAC GGTAGAATGA	3780
	CACCTAATAC CATTGTGATG ATAATTAATG CAATTGTTAT CCAAAACATT GTGTGACTTT	3840
10	GTTTCATGTCT CTTTCTTTTA GCAATCGACA CTTCATCAA TCCAACTACT GCAACACCAC	3900
	ACAGCATTFT CAATGTAAGC AACATATGAT TTGCCCCGCC ATTCATAAAT GACTGAATTA	3960
	ATATCCAAAA TCCTGAAATT AACGTCAACA GCATAAATAA GCGTAAAATC ATGTGCAACG	4020
15	GTTTGAAAAA TGGTGATCTG CCTTGATTTT TTGAAATGTT TAAGTATGTA GCGATAAATA	4080
	AAATAATCGC TAATACCCAA CTTAATATAT GTAAATGTAA CATACTGATT CCCCCACTT	4140
	TAATTATTTA TATTATTAAA TTAAAGCTTC TTGGGATTAA TACCCACTTG CTTGTAAATT	4200
20	AATCATGATT TGATTATACA CGAATATATA TTCTACCACA CTTCTATATT TGAGAGGAAG	4260
	AACATGACAT TTTATTCCTT ATTAGAATAT TGTGAATCTG CTGTAAAATA ATCAACTACT	4320
	TCTTAATATC AATATTTTAC TTTCATCTCA AAATGGTAAC ATTATAAATA ATTTATCTTT	4380
25	AACACCTTTT TAGAAAAGCA AGAAAAAACT AACCAATCTA TATAAAGACT GGTTAGCTTT	4440
	TTAAATGATA ATTATTTAGC GATATAAGTT GTCAGCGTTC CAATATTATC AATAGTCACT	4500
	TTAACTTCAT CACCTGGTTG TAAAAATTTA GGTGGTTGCA TACCTGCACC AACGCCTGCT	4560
30	GGTGTACCAG TTGCAATAAT ATCTCCCGGA TGTAGTGCAA CATATTTTGA AATTTCTTCT	4620
	ATTAATTCAT CAATTTTAAG AATCATTTTCG CCAGTGTTAC CATCTTGTCT AATTTCAATG	4680
35	TTAACTTTTG TAACAATATT TACATTTTCA GGTAATGGTA GTTCGTCTTT AGTAACGATA	4740
	TAAGGACCCA TTGGGCAACC GCCAGTTAAA CTTTTTGATA AAAATGCTTG ATCTTGTTC	4800
	CTTTGTGCTT TGCATCAGT GATATCGTTA ATAATTGTAT AGCCGTAAAC ATAATCTAAA	4860
40	GCTAATGCTT TTGGAATCTT TTCACCAGAC TTACCAATAA CAATACCTAA TTCACCTTCA	4920
	TAATCTAATT GATCAGTAAT ATCTTTATGA TTTGGAATTG TTGCATTATC TCCTGTTAAA	4980
	GATGACGCTG CTTTTGTAAA TACATATAAT TTTTCCACTT CATGATTTAA TTCGTTGCA	5040
45	TGATCTTTGT AATTTCTACC AAAAGCAATC ACATTATTCG GAGGTGTTAC TGGTGGTAAA	5100
	AATTCAATGT CATTAAATGA AATTTTATAG TCTTCAGCTT TGCCGCTATC TTCTGCTGCT	5160
	ACAACTGCTT TACGTACTTG TTCTTGAAAA TCTAAAGTAT GATTTTGTG TAAACCAGCT	5220
50	AACAATGTTT TAGGATGGAA ATCTCCTTCT GCAAAGTCAG CAAATACTTG TGTTAAATCC	5280
	CATACAGCAT CTTGCGTTT TACTTTAACG CCATATGAAG TTTTGTCAAT ATACTTGAAT	5340
55		



TTCGTTATCA AATAACAAAT AAATAAGTAA GACAATTTTG AAAATGAGTT GTGTTTCATTC 5460  
 TGCTACAAGG ACTTTGCACT TAATCGAAAT TATTTTTTAT TCTTTTGAAA ATCAAAATAC 5520  
 5 TATAGTTGCA ATGTACCAAA TTTGAAGAAG TATAAATAAC CTTTAACTTC TTTATTAAGA 5580  
 ATCGTTTGAA GCGTATTTTG ATAATATTTT ATCTGTATCT TATATTTAAT TTTTAATTGT 5640  
 10 GTACCAATTT CTTTCATCTGT CATCCCACGG CGACGATTAA ATGCATCGGT TTTATAGTCT 5700  
 ACAAATAAT GCACACCATC TTTAACAAAG ATTAAGTCAA TCATACCTTG AATAATTGAG 5760  
 ACG 5763

15 (2) INFORMATION FOR SEQ ID NO: 506:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 422 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 506:

25 CCATATGATT TTGTGCAATA ACTCTTTTTC TTTCTTCTTT TCGTAAAAAG TTGTACATCG 60  
 CTTTGTGTGTT TAAGAGACTA TTGTTTCTT TAGGTTTTTG AACTTCACTC AGTGTATTTT 120  
 TAGTTGTAA CACTAAAATT CCAACTGTTG TATCTTTGTA TCTAGCCATA ACTTTATTCA 180  
 30 GATGTTTGTC ATTTGTAATA ACTACGACAT AATTAAACAC TTCATAATAA TCATTAATTT 240  
 GATTATCTAA TCTATCCAGC TTATCTAATT CTGTTTAAAT CTCATAGACA ACGCCTTTGC 300  
 35 CGTTTAACAA TATAAATCA GCaATACTTT TCCCTATGGG CATCTCAGAA AGTGCAAGTAG 360  
 TTGTATTAAAT AGAATGTCGT CCTAGAAGGA GTkTATTAAG TATnGTGTTT TTGTAAAAAT 420  
 AT 422

40 (2) INFORMATION FOR SEQ ID NO: 507:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1188 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 507:

50 GCCTATGTAT TCTCAAAATA TTTATGTGAT ACGCAAAGGA GACATGGTTA TTCGACCAGC 60  
 ATTTGATGAT GACGATCAAA GAAACGGTAG TGAAATAATT CGGTTTGACA AAACGCGTAT 120

55

TTATCTTGGC AAGAAAGCAG AGACAAACCG CATTACTGGC ATTTCTAGTA AACCACCTAT 240  
 TTTACTAACA CCATTATTTT CAACTTATTT TTTCCCAACA CATTCTGACA GACAAAATGA 300  
 5 AAATATTTGG TTAAATATGC ATTATATCGA AAGTATTAAA GAATTAAAAA ATCGTAAATG 360  
 TAAAGTGACA TTTATTAATA ATGAATCAAT CATTCTTCAT GTTTCATACC ACAGTTTATG 420  
 GCATCAATAT AACAAATCCA TTTTTTACTA TTACATGGTA GATAACAAT CTCGCATGAT 480  
 10 ATCAAAAAAT CCCGACCAAC CAATAGATTA TAATAAAGCC ACATTGAATG TGTTTGAAGC 540  
 ATTGACACGC TATCTTTTAT TTGAAGATAA ATAAATTGTT TATTTTTTAA ATATGCGGAA 600  
 TGTTTTATAA ATATAGTGTA AATGTTCTGC ATATTTTTTT AAGGTATCTA TTGCAAATTA 660  
 15 ACTTAATCTT GTTATAATAA TATTTGTGCT TGaTATTCAA ACACATACAA ATTAATCCAC 720  
 AGTAGCTCAG TGGTAGAGCT ATCGGCTGTT AACCGATCGG TCGTAGGTTT GAGTCCTACC 780  
 TGTGGAGCCA TTGGAAACGT ACTCAAGTTG GCTGAAGAGG CGCCCCTGCT AAGGtGTAGG 840  
 TCGCGAAAg CGCGAGGGTT CGAATCCCTC CGTTTCCGTT ACTTGCTAAA ATGGTATATA 900  
 CCATTTTAnC TTTTTTGTTT ACTTATATAT AATGAATGAG AATTTCACTG TTCTTTTATA 960  
 25 TCAATTTTAA AATTCTAAAA ACCTTTCCTA GATAATCTTC TCTAAGAAAG GTTTTTTATAC 1020  
 TTGTTGAAC TATAATTAAT TTATTACATA GCAATATTTA CCTGTTTTTA ACTATAAAAT 1080  
 TATCACTACA TGAAATACGA TAATTCCGAt CTCTTAACTT CTCTGCaATT AATGtACTCA 1140  
 30 TTGgTTTCAT CGTATGATTC ATGTATAATA GCATTTkTTA AATAATTC 1188

## (2) INFORMATION FOR SEQ ID NO: 508:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 840 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 508:

CCCAACTTCG GTTATAAGAT CCCTCAAAGA TGATGAGGTT AATAGGTTTC AGGTGGAAGC 60  
 45 ATGGTGACAT GTGGAGCTGA CGAATACTAA TCGATCGAAG ACTTAATCAA AATAAATGTT 120  
 TTGCGAACAA AAtCACTTTT ACTTACTATC TAGTTTTGAA TGTATAAATT ACATTCATAT 180  
 GTCTGGTGAC TATAGCAAGG AGGTCACACC TGTTCCCATG CCGAACACAG AAGTTAAGCT 240  
 50 CCTTAGCGTC GATGGTAGTC GAACTTACGT TCCGCTAGAG TAGAACGTTG CCAGGCATAA 300  
 TATTAATCCA CAGTAGCTCA GTGGTAGAGC TATCGGCTGT TAACCGATCG GTCGTAGGTT 360

55

5 GGCGGTAACA CGGGTTCGAG TCCCGTAGgA GTCATACAAG CAGAAGTGAA ATATCGCTTC 480  
 TGTTTTTTTA TTACATATTT ATTGTTGAGG AAGGTTGTCC GAGCTGGCCG AAGaGCACGC 540  
 CTGGaAAGTG TGTAGGCGTC ACAAGCGTCT CAAGGGTTCG AATCCCTTAC CTTCCGTAAA 600  
 GGcGCTTAAA TTGGTTTTAC CCATTTTAAG CGCTATTTTT ATTTTGGACT CAATCCCTTG 660  
 10 ATATATCTGC ATTTGAGCTA TTATCCTCAT TTTTACACTT CTTATTTATT TATATCCATT 720  
 TAAATTTTT TAGCCACAAT GTGACTAATT TTTGaTGAAT AATCCTAATT TTAGtCACAA 780  
 GATTTTGAAG TTTAGTCACA AAACAAATCA TTCAGATTTT TTTCyATAAA TTTAGTTTCA 840

15 (2) INFORMATION FOR SEQ ID NO: 509:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1373 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 509:

25 TTCTATGAAA CTATCAGCTG TTTCTTTTTC CTTGGTAAAA ATCGTTTTAA CGCCATGATA 60  
 CATGTATTCT CTTTCAaCAC CCGATTGGTC ATAAGCATCG AAAACACAAA TAATTTTCATC 120  
 TGAAATAACT GCATTATAAT TTGCAATTGC ATCTATTAAT TGCAATTCTAG CTTCTTCTAA 180  
 30 ATTCTCTTTT GCAATGGCGC TTAGCGTTGG TGATTGTCCT ATCATATTGT ATCCATCAAT 240  
 GATTAAGTAA CGTTCTTTCA TTATATTTCT CCAACATCAT GTCTTTTTCG AAATACTTCG 300  
 TACATCATT AACTTGCTGC AACCgaAGCA TTCAAAGTGT TTACATGTCC AACCATTGGA 360  
 35 ATCTTAATAT AAAAATCGCA TTTATCACTT ACTAGGCGAC TCATACCCTG TCCTTCGCTA 420  
 CCAATTACAA TAGCCAATGA CATGTCCGCT TCTAGATTTC TATAATCTGT TGCAATTATTA 480  
 40 GCTTCAGTGC CAGCTACCCA AAAGCCATTA TCTTTTAGTT CATCGATAGT TTTAGCTAAA 540  
 TTTGTCACTC GAATAACTGG TACATGTTCA ATTGCACCTG TTGAGGCTTT TGCAACTGTT 600  
 TGCCTTAGTG TAACTGAACG ACGTTTAGGA ATAATAACAC CATCAACTCC CGTTGCATCG 660  
 45 GCTGTTCTTA AAATTGATCC CAAGTTATGT GGGTCTTCTA AGCCGTCTAA TATAAGTACT 720  
 GTCAATAAAC CTTCTTTTTC TTTTGCTGT TTTAAAAATT GATCGAAGTC AGCATATTCA 780  
 TATGGTGCAA TAAGCGCTGC AACACCCTGA TGTGGTGCAT TTGCTAAAAA ATCTAATTTA 840  
 50 GATTTTGGTA CAGTTTGAAC AATGATTTTT TGATCTTTTG CATTTTTTAA AATTTCAATTA 900  
 ATTTGTTGCT TTTAATACC TTCTTGAATC AATATCTTAT TTAYCGGATG CCCAGTAATA 960

55

TTCGTTTACT ATTGTTATTA TTTTATTTAA TAATGCCTCT AATCGTTCTT CTCTTTTTTC 1080  
 TAAATAaAGA AAACCTATCA CTGCTTCTAT CkCTGAACTT TTACGATATG TTTGAACATC 1140  
 5 AGTGTTTTTA GCTTTAGTAT GACTTTTCGC GTTACGCCCT CGCTTCAAAA TATCCATTTC 1200  
 TTCGTCTGTA AACCATTCTT GCTCCATTAA ATATTCTAAC GTTTGCGCCT GACTTTTGGC 1260  
 AGATACATAT TTTTtagACA TTTGATGTAG TTTATTAGGC TTACTTTTAA GCTTTAAaAC 1320  
 10 GATATAGGTA CGTACATATT GATCTAaGAC TCGGTcnCCC ATATATGCTA AGG 1373

## (2) INFORMATION FOR SEQ ID NO: 510:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 717 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 510:

TCAAGTGCCA ACTACACAAC TGATTTACAT TCTTTAGGTC AATATGTACA AGAAGGCCGT 60  
 25 CGCTTCTTAT TCGAAACAGT TGTAaAGTA AATCATCCTA AATATGATAT TACTATTGAA 120  
 AAAGATAGTG ATGATCTAGA CGGATTAAAT TATTTAGCTG GTAAaACAAT CGACGAaGTT 180  
 AACACAAAAG CATTcGAAGG TACATTATTA GCGCATACTG ATGGTGGTGT TCCTAACATG 240  
 30 GTAGTGAACA TTCCACAATT AGATGAAGAA ACTTTcGGTT ACGTCGTATA CTTCTTCGAA 300  
 CTTGCTTG TG CAATGAGTGG ATACCAATTA GCGGTAAATC CATTTAACCA ACCTGGTGTA 360  
 GAAGCATATA AACAAACAT GTTCGCATTA tTAGGTAAAC CTGGTTTTGA AGaCTTGAAA 420  
 35 AAAGrATTAG rAGAACGTTT ATAAaATACA TTACTTCAAA GATTAGTGAA GTTTGAAAAG 480  
 ATAGAACTAG ACGTTAACTA TTTAAAGCAT ATTTTCGAGG TTGTCATTAC AAATGTAAAA 540  
 40 ATGTAATGAC AACCTCGTTT TTATTtATAT GCAAGAACTA GGTTACTAGC TAaTG TGACA 600  
 AGATGTTAAG AGAAAATTAA aGATAAAATA ATATCTGCCT TACAATAATA TTGTTATACT 660  
 ACTAGnGACT GATTTATTAG CATGATTACA TGTTAATGTT TCTTTACTTA GTAATTA 717

45

## (2) INFORMATION FOR SEQ ID NO: 511:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2700 base pairs  
 (B) TYPE: nucleic acid  
 50 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

	AATCTAATTT TTCCTCCTAA TCTCGGATTG GATAGTGGTC AATTTTTCTG GCCCGCCATT	60
	TTAGCGTTTG TTCTAACTGG GATTGGTTTA CCATTATTAG GTGTGATTGT AGGTGCACTT	120
5	GATAAAGAAG GATATATTGG CGCATTAAAT AAAATTTTAC CTAAATTTTC AATATTGTTC	180
	TTAATCATCA TTTATTTGAC TATAGGACCA CTTTTTGCAA TACCTAGAAC TGCATCTACA	240
10	TCTTTTGAAA TGACAATTAC ACCAATTATA CATAGCAATA GTAGTATCGC TTTATTTATA	300
	TTTACGATTA TCTACTTCAT AGTCGTTTTG TATATTGTGT TAAATCCATC TAAGTTAATC	360
	GATCGTATTG GTTCATTATT AACACCATTA TTATTGATTA CTATTTTAGC GATGATTATT	420
15	AAAGGATACT TAGACTTTAG CGGTAATAGT GCTGGAAGG GCAATGAAGC ACTATATCAT	480
	TCTAATTTTT CAAGTTTTGC TGAAGGCTTT ACACAAGGCT ATTTAACAAT GGATGCCATT	540
	GCAGCAATTG CTTTTTCAAT GATTGTTGTT AATGCAGTAA AACTAACAGG CATTACTAAA	600
20	ACAAATCAAA TATTCAAACA AACTTTGACT GCTGGTTTAA TTGCAGCCGT AGCTTTAATT	660
	TTCATATATA TTTCATTAGG TTATATTGGT AATCATATGC CAGTAAGTGA CATGACGTTA	720
	GATCAATTGA AATCCAAAGA TCGAAACATT GGGACATATT TATTAACGAC AATGGCTTCA	780
25	ACAGGATTTG GTTCATTTCG AAAATATTTA TTGGGCATCA TTGTGGCGCT GGCATGTCTA	840
	ACTACAGCAT GCGGGCTTAT TGTTGCAGTT TCTGAATATT TCCATAGAAT CGTACCTAAA	900
30	GTATCATACA AAGCATTTGT ATTAGTTTTT ATTTTAATGA GTTTTATTAT TGCTAACCAA	960
	GGTTTAAATG CTGTTATCTC AATGTCAATT CCGGTATTAA GCATTGTATA CCCAGTAGCA	1020
	ATAACTGTTG TATTATTAAT TTTAATTGCC AAATTCATAC CGACAAAACG CATTTACAA	1080
35	CAAATTCAG TTATTATCGT ATTTATATTG TCGATTTTCA GTGTATTAG TAAGTTAGGT	1140
	TGGcTGAAAA TTAACTTTAT AGAATCATTG CCTCTAAGAG CGTATTCTTT AGAGTGGkTC	1200
	CCAGTAGCAA TTATTGCAAC GATATTAGGC TATCTAGTCG GCATATTTGT AAAACAAGAT	1260
40	CCAATTAAAT ATCAACAGGA ATAACGAATA ATATAAAAGA GGTGGGACA TAAATCCCTA	1320
	AAAAAACAGC AGTAAGATAA TTTTCAATTA GAAATATCT TACTGCTGTT CTCTATTTAT	1380
	ACAATACTTC GTATTGAATG GCTTCGCTTT CCTAGGTGTC CGTCTCAGCC TCGGTCTTCG	1440
45	ACTGGCACTG CTCCCTCAGG AGTCTCGCCA TTAATACTAC GTATTAACGT GTAATTTTAC	1500
	TTTGAAATAC TTTAAAAAAA TAAGACACTT TGCCCAACTT GCACATAAAT GTAAAATTCA	1560
50	ATAAAATAAA TTTTCTGTGT TGGATCCCTT CGTATAATTT AATAAATACT ACTAAACTAA	1620
	ATTAACGAGG TGCCTTATGT ATAAAAATTA TAACATGACC CAACTTACAC TACCAATAGA	1680
55	AACTTCTGTT AGAATTCCTC AAAATGATAA TACGCGATAT GTTAATGAAA TTGTTGAAAC	1740

5 AAGATATGCA TACCGTAATG ATAGATATAG TTTTAAACGT GACTTCAAGC TATATGAATG 1860  
 TGATGACTGT TCATCATGTT CTTTGAGACA TCAATGCATG AAGCCAAATT CGAAATCCAA 1920  
 TAAGAAAATT ATGAAGAATT ATAATTGGGA ATACTTTAAA GyCCAAATTA ATCAAAAGCT 1980  
 TTCTGAACCA GAAACGAAAA AAATCTATAG TCAAAGAAAA ATTGATGTAG AGCCTGTTTT 2040  
 10 TGGATTTATG AAGGCTATTT TGGGTTTCAC TCGAATGTCA GTTCGAGGAA TAAATAAAGT 2100  
 TAAACGAGAG CTAGGTTTGT TATTAATGGC ACTTAATATA AGGAAAATAG CAGCTCAACG 2160  
 AGCTGTACAT TATAAAATAC ATATCAAAAA AGCTGATTTT TATCAAATAA TTAATAGAAA 2220  
 15 TCAGCTTTTT TACATTGCCT AAGAATTTAA TGTCCCAAGC CCTTTTTATC GAATAACTTA 2280  
 TTGTAAACCT TGTCTTTCTT GGTATTGTT TTCGTTATTT TTTTCGTGTT TTTGTTTCCA 2340  
 CTCTTTTTGA GTCATTACAT CGTCAACTTG CATGTAACT TCAACTACAT CTAAACCACT 2400  
 20 AATATATTTA ACTTGTTCTT TAACTAATTC AGTTACTTTA CGGAAGATTT TTGGTGCTGA 2460  
 TTCACCATAT TCTAAGATTA CTTTAAAGTC TACAGCAGCT TGTTTTTCAC CAACTTCAAC 2520  
 AGATACACCT TGAGTAACAT TGTGCGCACT TGAGAATGCA TTAGTGAATG TATCAGTTAA 2580  
 25 GCCACCTTTC ATATCTAAGA TACCTTTAAC TTCACGTGCA GCGATACCAG CAATTTTTTC 2640  
 AACAACTTCA TCAGAGAATG TTAATTTGnT TTTGAATTGA GGCTCCTGAT TTTnGTTCTnTG 2700

30 (2) INFORMATION FOR SEQ ID NO: 512:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1135 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 512:

40 ATCAAATACA ATAATTAAAA TAaTGGCTAA TACaCCTAAa ATaATAATcG AGTTaGAGTA 60  
 GCCTAAGCGA yCACCcTAAC aGTArACATT CTAGGCATAT ATTTaTCTTT AGACATTGAA 120  
 GCCGCTAACA TCGGAAAAGC TGTAaATCCA GTGTTGGCTG CTAATACCAa AATCATCACG 180  
 45 GTCGTTGCCT GTACAAAATA GAAGGCAGCA TTATCACCAa ATATTTGCAT AGCTAATTGT 240  
 GATAAAACTG TCGTTTCCGT TTGTGGcAAA ATTCCATAAA CATATGCTAA ACCAACGATA 300  
 CCAACTAATA AAAACGCTAA AATTGAACCC ATAGCAATTA ACGTTTTTAC AGCATTTTTA 360  
 50 GCACTTGGCT CTCTAAAATT GGTtACCGCA TTTGAAATAG CTTCAACACC CGTTAACGAT 420  
 GAAGCCCCTG ATGAAAATGC TTTTAATAsC AAGAATAATG TTACTCCAGG AACCGCAGTT 480

55

AATATTAATA TCACTAACCC TATAATGAAA AGATATACTG GATAGGATAA TACGGTGGCA 600  
 GATTCAGTTA AACCACGTAA ATTTAATATT AAAATAAAAA GTACAAGTAA ACATGCAATC 660  
 5 AGTCTCTTAT GCCCATATAA ACTTGGGAAT GCAGCAACAA ATGCATCAGC ACCAGATGAT 720  
 ATmCTAACAG CGACAGTCAG TATGTAATCG ACTAATAATG AGCCTCCTGC AAGCAATCCC 780  
 CATTITTCTC CTAAATTGGT CTTGGACACC ATATACGCGC CGCCACCTTT AGGATATGCA 840  
 10 TAAATAATTT GCCTATAAGA CATAATTAAA GCAGCTAATA AGATCAGAAC AGCACCTGCA 900  
 ATCGGTAAAG TATACCAAGT TGCAACTGCA CCCACTACTG ATAGTGTAAT CAGTATTGT 960  
 TCGGGACCGT AAGCCACTGA AGATAATGCA TCCGACGAAA GAATCGCTAA CCCTTTAAAC 1020  
 15 TTCGATATCT TTTCGTCTTT TAGTTCTCTG TTTTTTTTAG GTTGCCCTAT AATAAGCTT 1080  
 TTAAATTGAT kGACATAATc TCCTATTCCT TTTTATAGTT TtngATGGAA ATCAC 1135

20 (2) INFORMATION FOR SEQ ID NO: 513:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 918 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 513:

TTATGGTAAA ACTTTAACAC AAGCATTACA TAAATCGCAA TCATTTATGT GGCAAAAACG 60  
 TTGAATTTAA GCCATTTCTA TTCATTTTAC GTAATTATTA GCCGTATATG TGTAATAATA 120  
 35 CACATTTTAT TCAGATTTTT TATCGCGCTC CATTAAATCT TTTACGCATT CTTTTACTGA 180  
 GATATTTTCA AATAATACTC TATATAATGC ATTTGTAATT GGCATATCCA CATTTTTTTTC 240  
 TTTAGCTAAA TGATAAACTG ATTTAGTTGT ATAAATACCT TCAACAACCA TATTCATTTT 300  
 40 AGATAATGCT TGATCCATTG ATTCACCTTG TCCAAGTTTA TATCCTAATG TGAAATTCGG 360  
 AGAATGTGTT GATGTGCAAG TAACGATTAA GTCACCGATA CCACCTAAAC CTAGAAATGT 420  
 CATAGGATCG GCACCTAACT TTTCACCTAA TCTACTAATT TCCGCTAAGC CACGAGTCAT 480  
 45 TAATGCAGCT TTTCATTAT CACCGTAGCC AATTCCAGCT ACGATACCAC TTGCTACTGC 540  
 GATGATATTT TTCAATGCAC CACCAAGTTC AACACCAATC AAGTCATCAT TCGTGACAC 600  
 ACGCAAATAA TCATTCATAA ATAAATCTTG CGTTAATTTA CTTACACTTT TATCTTTTGA 660  
 50 TGAAGCAGCA ACTGTAGTTG GTTGCTTGAC TACAACCTCT TCCGCATGAC TTGGCCCTGA 720  
 CAACACGCCA ATACCTGCAT TATATTCAGG TGAAATAGAA TCTTCAATCA TTTCTGACAC 780

55

CAGCTTATCA TTAATTTGAG AAGCAACTTC TCGCATTGCT TTAGTAGGTA AAGCCATTAA 900  
GTAAATATCT GCAAATTG 918

(2) INFORMATION FOR SEQ ID NO: 514:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 587 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 514:

CAATACTCGG TATTATAGAT AGTCCTACTA CATTTATATG GGTTCGTATC ATTACATTAA 60  
TTGCACATCA ACTTGAAGGT AATATCATT A CACCAAATGT AATGGGTAAA TCTTTAAGTA 120  
TCCATCCTTT AACAAATTATC GTTGTTATTT TAGCAGCAGG AGATTTAGGT GGCTTTACAT 180  
TAGTTCTGAT TGCAGTGCCA TTATATGCTG TACTTAAAAC GGTTCGTTAGT AATATTTTCA 240  
AATACCGCCA ACGCATTATT GATAAAGCAA ACAGTAATGT TAAGGACTAA TTCTGTGGAT 300  
GTCTTTTAAG AATATAAGAT ACTATCGCAT CAAAAGTTGA AACTACAGCT TTTGAGGCGA 360  
TTTTTTTGTG CATAAAAAAT CAGTCMAATG AAATATCAAA TAATTTTCCA TCAGTCCGAT 420  
TATTATAAAA GCAAAAAAGC TTTGCTCACA TATATAATAA CGTGAACAAA GCTGTTGAAT 480  
GATATTATTT AATTGCGTGG AATCCGCTAT CTACATGAAT ATTTTCACCT GTAACGCCAC 540  
TTGATAAATC ACTTAATAAG TAAGCCGCAG TTTTACCTAC TTCTACT 587

(2) INFORMATION FOR SEQ ID NO: 515:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 812 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 515:

GGTTCATTCC AAAAAAGTAC GCGATAATTA GCGAAGAAGA ATTTGAAAAT TTAAATGTTG 60  
TTAAACCAAA TAAAAATAAT GTTTTCTGGT CAGTTATAGG AAGTTCGGTT TTGTTTGGAG 120  
TTACTTTAAG GAAATACATA CATGTTTTTG ATGTTCAATT AGATAAGCTA GTTGTAATGA 180  
TATTGTGTGC TCTCGCTTTA ATTTGTGTGA TAGTTTTTTA TTTTAACTTA AATAGAAAGC 240  
TTAAGTTAAA AGTGTTTGAT ACAAATATTG AAAAAAATAA GAGAGTTATA TAAAwACCAA 300



5 TTTCAATTAAT TGCCCTTATG ACAATCGAAC CTCAAAATAT AATAATATTT ATTTATTGGA 420  
 TTATGATGAC AATGCTTTTC TTTTGTGTTAA ATATGACTTC GATAGGTAAT GAAAAAGTTC 480  
 GCGTTATAAT GAAAAATAAT TGATTACATT TAAAATATTC TAAATGTTGT CGACACAATC 540  
 CTTTTAAGAC GCTAGTAGAA TTTAAATGAC TTCTAATGTA TATGAAAGTG TATCAATATA 600  
 AAACCAATTG AAAAGAAGTG GAGACATTGC TTTGTGAAAC TGAAAATATT AATAAGAATC 660  
 10 CCAAATATAG AATTATCAAA TACAAAGATG AATATTTGAT GATTGATTTA GTAAGTACAT 720  
 GGTTAGCACT CTTTTTCCCA ATGATTAATT GGCTGATTCC AAAAAAGTAC GTCAAAATCA 780  
 15 GCGAAAAAGA TTTTGAACT TTAAACATTG TG 812

(2) INFORMATION FOR SEQ ID NO: 516:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 526 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 516:

GTTCTATTAT TTCTTCCAAT GGTAATGGAT TTTCGTAAAA TGATTGATAT AAGTTGATAA 60  
 TCTCAATAAC TTITAATTGA TCTGGAAACA TCGTTTTTTG GAACATTATA CTGCGATTTT 120  
 30 CACTTTGTAA TAACTTGTCT TTATCAAATA TCTCACCAGA ATTAGCATTACATTACCAA 180  
 TTAATATATC AATTAACGTT GACTTTCCAG CACCATTTTT TCCAATTAAA GCGATACATT 240  
 35 TACCTTGTTT AATATCGAAC GAAATATTTT TTAGAACACA TCTTTTATTA AATGACTTGT 300  
 TGATATTAGA TATTTGAATC AATTTAATCA CCTCTATTTT TTTCTTAATT TAATATTAGT 360  
 AAATTTATTA GATTTAAAT AGAACAACCTT GTCATAGATT TGAAATGACA AATGTCATTA 420  
 40 TTAGATktAC ATAATATATT TATCGTtATT TTAATTTTGG GCAAAATAAA AAGAGCCTCT 480  
 ATAATCGrGc TCCTTACAAA TaaATTATAA AAttGGCGAA cTAAAT 526

(2) INFORMATION FOR SEQ ID NO: 517:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4544 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 517:

55

	GGAACACCTA AAGAAAGAAA TGATGCATTA AACACAGAGG CTGATATCTA TGTAACCAAT	120
	AAAGAAAATA CTAAATGGTT ATGCGATCAA TATAAAAAAG AATGGCCATT TGATATGGTT	180
5	GTAATTGATG AACTGTCTAC ATTTAAAAGT CCTAAGAGTC AAAGGTTTAA ATCTATTAAA	240
	AAGAAATTAC CACTCATTA TAGATTTATA GGATTAACAG GAACACCTAG CCCAAATAGT	300
10	TTACAGGATT TATGGGCTCA AGTTTATTTG ATAGACAGAG GCGAAAGACT TGAGTCTTCA	360
	TTCAGTCGTT ATCGAGAAAG GTACTTTAAA CCAACACATC AAGTTAGCGA ACATGTTTTT	420
	AACTGGGAGC TAAGAGACGG ATCTGAAGAA AAGATATATG AACGAATAGA AGATATATGT	480
15	TTAAGCATGA AAGCGAAAGA TTATCTAGAT ATGCCTGACA GAGTTGATAC TAAACAAACA	540
	GTAGTCTTAT CTGAAAAAGA AAGAAAAGTA TATGCAGAAT TAGAAAAAAA CTATATTTTA	600
	GAATCGGAAG AAGAAGGAAC AGTTGTAGCT CAGAATGGGG CATCATTAAG TCAAAAACCTA	660
20	CTTCAACTAT CTAACGGTGC AGTTTATACA GATGATGAAG ATGTAAGACT TATACATGAT	720
	AAGAAGTTAG ATAAGTTAGA GGAAATTATA GAGGAGTCTC AAGGCCAACC AATATTATTG	780
	TTTTATAACT TCaAACATGA TAAAGAAAGA ATACTTCaAA GGTtTAAGGa AGCAACCACA	840
25	TTaGAGGATT CAAACTATAA AGAACGTTGG AATAGTGGAG ACATTAAGCT GCTTATAGCA	900
	CATCCAGCAA GTGCAGGGCA TGGATTAAAC TTACAACAAG GTGGGCACAT TATTGTTTGG	960
30	TTTGGACTION CATGGTCATT GGAATTATAC CAACAAGCAA ATGCAAGATT ATATAGACAA	1020
	GGACAAAATC ATACGACTAT TATTCATCAC ATTATGACCG ATAACACAAT AGATCAAAGA	1080
	GTATATAAAG CTTTACAAAA TAAAGAACTA ACGCAAGAAG AATTGATGAA AGCTATTAAA	1140
35	GCAAGAATAG CTAAGCATAA GTAATGGAGG TATAAGATGG GAAAGGCATC ATACGATATT	1200
	AAGCCAGGTA CATTTAAATA TATTGAGTCA GAGATATATA ACCTACAAGA GAACAAGAAA	1260
	GAGATAAATA GATTGAGAAT GGAGATACTT AACCCAACGA AAGAGCTAGA CACTAACATT	1320
40	GTGTATGGAC CGTTGCAAAA AGGTGAACCA GTTAGAACAA CTGAACTAAT GGCAACAAGG	1380
	TTATTGACTA ATAAGATGTT ACGAAACCTA GAAGAAATGG TCGAAGCAGT TGAAAGTGAA	1440
	TACTTAAAGT TACCTGAAGA TCATAAGAAA GTAATTAGGC TAAAGTATTG GAATAGAGAT	1500
45	AAGAAGCTAA AGATAGAGCA AATAGGAGAT GCATGTCACA TGCATCGTAA TACAGTTACT	1560
	ACTATACGAA AGAACTTTGT TAAAGCGGTA cGwtATCATG CAGGTATCAA ATAACATTGT	1620
50	GCAAGATTG TGCAAAAGGC CTACAAATCT GTAGTAATAT GATAGTATCG GAAAGATGTA	1680
	TAAAGTTATC TAAAAGTTAT ACGACACAAG TACACGAGGC ACATCGCTAT GCGtGTGTCT	1740
55	TTTGTTATGC AATCAAAGAG GTGTAAGAGA TGACCAAGCA TAATAACATT TATAAGCATG	1800

	AGATAGCATT AGATAGGGAT AATCATCTTT GTCAAATGTG TTTACGTGAA GACATAGTAA	1920
	CAGATGCAAA CATAGTGCAT CATATTATTT ATGTTGATGA AGATTTTAAT AAAGCTTTAG	1980
5	ACTTAGATAA TTTGATGTCA GTTTGTTATA GCTGTCATAA CAAAATTCAT GCAAATGATA	2040
	ATGACAAAAG TAATCTTAAG AAAATTAGAG TATTAAAAAT TTAAATAAAA AAATAATTTA	2100
	TTTTTATAGC CCCCTACCCA TCGGCTTAAA ATGTTTMTTC GACGGGTACC GCGGGGGGCC	2160
10	CTTCGCTTGC AACGCGGATA AACTTTTATG AAAGGGGGTC TTTATATGAA ATTAACAAAA	2220
	AAACAGCTGA AAGAATATAT AGAGGATTAT AAAAAATCTG ATGACATATT AATTAATTTG	2280
15	TATATAGAAA CGTATGAATT TTATTGTCCG TTAAGAGATG AACTTAAAAA TAGTGATTTG	2340
	ATGATAGAGC ATACAAACAA GGCTGGTGCG AGCAATATTG TTAAGAATCC ATTAAGCATA	2400
	GAACTGACAA AAACAGTTCA AACACTAAAT AACTTACTCA AGTCTATGGG TTAACTGCA	2460
20	GCACAAAGAA AAaAGATAGT TCAAGAAGAA GGTGGATTCG GTGACTATTA AAGTTTTAAA	2520
	TGAACCTTCA CCAAACTAT TAACAACATG GTATGCAGAG CAAGTCACTC AAGGGAAAAT	2580
	AAAAACAAGC AAATATGTTA AAAAAGAATG TGAGAGACAC CTTAGATATC TAGAAAATGG	2640
25	AGGTAAATGG GTATTTGATG AAGAATTAGC GCACCGTCCT ATTCGATTCA TAGAAAAGTT	2700
	TTGTAAACCT TCCAAAGGAT CTAAACGTCA ACTTGTATTA CAACCATGGC AACATTTTAT	2760
	TATTGGCAGT TTGTTTGGTT GGGTTCATAA AGAAACAAAA CTGCGCAGGT TTAmAGAAGC	2820
30	TTTGATATTT ATGGGGCGAA AAAATGGTAA AACAACTACT ATATCTGGTG TTGCTAACTA	2880
	TGCTGTTTCT CAAGATGGAG AAAACGGCGC TGAAATCCAT CTTTTAGCAA ACGTAATGAA	2940
35	ACAAGCTAGG ATTCTATTTG ATGAATCTAA GGCGATGATT AAAGCTAGCC CAAAGCTTAG	3000
	AGAAAATTTT AGACCTTTGA GAGATGAAAT TCATTACGAT GCAACTATAT CTAAAATTAT	3060
	GCCACAGGCT TCAGACAGTG ATAAGTTGGA TGGTTTAAAT ACACATATGG GCATTTTTGA	3120
40	TGAAATTCAT GAATTTAAAG ATTATAAATT GATTTCAGTT ATAAAAAACT CAAGAGCGGC	3180
	AAGGTTACAA CCCCTTCTTA TCTACATTAC GACAGCAGGG TACCAACTAG ATGGACCACT	3240
	TGTTAATATG GTAGAAGCGG GAAGAGACAC CTTAGATCGA ATCATCGAAG ATGAAAGAAC	3300
45	TTTTTACTAT TTAGCTTCTC TCGATGATGA CGATGATATA AATGATTCGT CGAATTGGAT	3360
	TAAAGCAAAT CCTAACCTAG GTGTTTCTAT CGATTTAGCT GAAATGAAAG AAGACTGGGA	3420
	AAAGGCTAAG AGAACACCAG ATGAACGTGG AGATTTTATA ACCAAAAGGT TTAACATCTT	3480
50	TGCTAATAAT GATGAGATGA GTTTTATTGA TTATCCAACA CTTCAAAAAA ATAATGACAT	3540
	TATTTCTTA GATGAGTTGG AAGGTAGACC ATGTACTATA GGTTATGATT TATCAGAAAC	3600

55

AACACATTCT TGGATTCTTA AGCATAAAGT TGAATATTCT AACGAAAAGA TACCCTATAT 3720  
 AGAATGGGAA GAAGACGGAT TACTAACAAT ACAAGATAAT CCTTATATAG ACTACCAAGA 3780  
 5 TGTTTTAAAT TGGATAATAA AGATGAATGA GCATTATGTT GTCGAAAAAA TCACTTATGA 3840  
 TAGGGCGAAT GCTTTTAAAT TAAATCAAGA GTTAAAGAAT TATGGCTTTG AAACAGAAGA 3900  
 AAcwAGACAA GGGGCTTTGA CCTTGAGCCC TGCaTTGAAG GATCTAAAAG AAATGTTTTT 3960  
 10 AGATGGGAAA ATAATATTTA ATAATAATCC TTAAATGAAA TGGTATATCA ATAATGTTCA 4020  
 GCTGAAACTA GACAGAAATG GGAAGTGGCT GCCATCTAAA CAAAGCAGAT ATCGTAAAAT 4080  
 15 AGATGGTTTT GCAGCATTTT TAAACACATA TACAGATATT ATGAATAAAG TTGTTTCTGA 4140  
 CAAGGGTGAA GGAAACATAG mATTTATTAG TATTArAGAT ATAATGCGTT AAGGAGGTGA 4200  
 ATGTTATCGC AAAAGAGAAT ATTGTCACAC GCATAAAGAA AAAATTGATA GACAATTGGA 4260  
 20 TTGaTCAGTC AGCTTCTAAG CTTTATGACT TTAGCCCATG GAAAAATAAA TCTTTTTGGG 4320  
 GTGTAATCAA TAATACGCTT GAACTAATG AAACGATATT TTCAGCTATT ACnAAGTTAT 4380  
 CTAATTCGAT GGCTAGTTTG CCCTGAAAA TGTATGAAGA TTATAAAGTA GTTAATACAG 4440  
 25 AAGTATCTGA TnACTTACA GTGTCACCGA ATAATTCTCT GAGCAGTTTT GATTTTATTA 4500  
 ATCAAATTGA AACAATCAGA AATGAAAAAG GTAATGCATA TGTG 4544

(2) INFORMATION FOR SEQ ID NO: 518:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 881 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 518:

40 CATGGCAGAA TATTGAAGCT GTAAAAAAG GACATGTAAT TTCATATAAA GCAGAAGATT 60  
 ATTGGTTCAC AGATCCTATT ACATTAGAAC ATTTGAGAAG TAAATTAAAA AAAGAAATTT 120  
 TAAATAAAAA ATAATAGAAA TAAGTTGTAA AAATTTTCTT ATGCATTGGT ACTAATGTTT 180  
 45 TTAAGGAGTG ATTAAATGAA GCAACTGGTT GGAATCCCG AATCAATGTT AATTCCTTTG 240  
 ATAGCTCGAG CAAAAGAGTA CGAAAACGAA AAACCAATAA TAAAAGACGC ACTATCAAAA 300  
 AAAATATTTG ATGGTTTGA TAGATATGTAC AAAAATGTTA CATGTGATGA CATGTCTCAA 360  
 50 ATTGGAATTA GTATACGTAC TGTGATAATA GATTGTGTTA CTAAGAGGCT TATCAAGGAT 420  
 AATAAAGATT TAATCGTGGT CAATATAGGT TGTGGCTTAG ATACAAGGTT TCAAAGATTT 480

55

EP 0 786 519 A2

ACATTTTTTA AAGAAAGTAA TAGTTATAAG ATGATATCTA AATCTATGCT AGATTACAGT 600  
 TGGATTGATG ATGTCAAAAA TTATAAATTT TTTAATAGTA AGTCAGATAT ATTGTTTATT 660  
 5 ATTGAAGGTG TATTGaTGTA TTTTGATGAG AGTGTAAATGA CTCAATTATT GGACACTATT 720  
 ATCAAAAAGA TGGGAGATCA TAATTTGACA TTTGCGATTg AATTTTGCTC aAAAAACAATT 780  
 GCGAATAATA CmAAGAGACA TCAATCGGTA TCCAAGTTAT CCTCACCACC TGTTTTTAAA 840  
 10 TATGGGTACC ATGATTTAnA AAAATTGGAT GAnATTTACC C 881

(2) INFORMATION FOR SEQ ID NO: 519:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3122 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 519:

TGAATaAAAA TATATTAATA GATAAACACA AATGTGTcCA AATACCCCTA GAGGTATTTG 60  
 25 AChAGTTCCA TCCAACGTGT TAAATACCC CTACAGGTAT TTTTAGGGAG GTTATTATGA 120  
 AACAAATACGG AGAAAAgTTT ATCGATGAAT TTAGTAAAGC AGAATTGGAA AAAC TAGCCA 180  
 AGCAAGGGCA ATTAATTGAC GTTAGAACAG AAGAGGAGTA TGCATTAGGA CATATCAATG 240  
 30 GTTCCATACT TCATCCTGTT GATGAGATTG AGTCATTCAA TAAAGAAAAA AATAAACCT 300  
 ATTATGTAAT CTGTAGAAGT GGTAACAGAA GTGCTAATGC TAGTAAATAT TTAGCTAAAC 360  
 AAGGTTATAA CGTTATAAAT CTTGATGGTG GTTATAAAGC TTATGAAGAA GAAAACGATA 420  
 35 GTTATGATAC ACAAGAAGAA TATAAAAGTA TAGAAATTAA AGCAGATCGT AAACAATTTA 480  
 ACTATCGTGG TCTTCAATGT CCAGGGCCAA TTGTAAAAAT TAGTCAAGAA ATGAAGAATA 540  
 40 TTGAAGTAGG TGACCAAATT GAAGTCAAAG TCACAGACCC TGGATTCCCT AGTGACATTA 600  
 AAAGTTGGGT GAAACAAACA AGGCATACTT TAGTTAAGCT TGATGAAAAT AACAAATGGAA 660  
 TTAATGCGAT TATTCAAAAA GAAAAAGCAA AAGATTTAGA TATAAATTAT TCTGCTAAAG 720  
 45 GTACTACAAT TGTATTATTT AGTGGAGAAT TAGACAAAGC TGTAGCAGCG TTGATTATTG 780  
 CAAATGGTGC TAGAGCTGCT GAAAAAGATG TAACTATCTT CTTTACTTTT TGGGGGCTTA 840  
 ATGCATTAAA AAAAGTGCAA ACAGTTAATG TTAAAAAGCA AGGTATTGCA AAAATGTTTG 900  
 50 ATTTAATGTT GCCCAAAAAG AATATACGAA TGCCTCTTTC CAAAATGAAT ATGTTTGTTG 960  
 TAGGAAATAT GATGATGCGC TACGTAATGA AAAAGAAAAA TGTTGATTCA TTACCAACAC 1020

55

5 TCATGGGTAT TCAGAAAGAA GAACTTAGAG ATGAAGTTGA GTACGGTGGT GTAGGCACTT 1140  
 ATATTGGTGC TACTGAAAAT GCGAATCATA ATTTATTTAT CTAATTAAAT CTATTAATAA 1200  
 AAGGAGTTGT TATCATGTTT TTAAACAGT TTTACGATAA TCATTTATCT CAAGCATCAT 1260  
 ATTTAGTGGG TTGTCAACGT ACAGGAGAGG CAATAATAAT AGACCCTGTT CGTGATTTAT 1320  
 10 CGAAATATAT AGAAGTTGCA GATTCTGAAG GTTTAACAAT TACACAAGCT ACAGAAACAC 1380  
 ATATTCATGC TGATTTTGCT TCAGGAATTC GTGATGTGGC TAAACGCTTA AATGCAAATA 1440  
 TATATGTGTC TGGCGAAGGT GAAGATGCAT TAGGGTATAA AAATATGCCA TCAAAAACAC 1500  
 15 AATTTGTAA ACATGGAGAT ATCATTCAAG TAGGCAATGT TAAATTAGAA GTTCTGCATA 1560  
 CTCCAGGACA CACGCCTGAA AGTATTAGCT TTTTACTCAC TGATTTAGGT GGTGGTTCAA 1620  
 GTGTTCCGAT GGGATTATTT AGTGGTGA CT TATTTTGT TGGTGATATA GGTAGACCTG 1680  
 20 ATTTATAGTA AAAATCTGTT CAATAAAAG GTTCTACAGA AATTAGCGCG AAACAAATGT 1740  
 ATGAGTCCGT TCAAAATATT AAAAATTTAC CAGACTATGT TCAAATCTGG CCGGGTCATG 1800  
 GTGCTGGAAG CCCTTGTTGGT AAAGCATTAG GTGCCATACC TATATCTACA ATAGGTTATG 1860  
 25 AGAAAATTAA TAACTGGGCA TTTAATGAAA TTGATGAGAC TAAATTTATT GAATCATTAA 1920  
 CATCAAATCA ACCAGCACCA CCGCATCATT TTGCACAAAT GAAACAAGTT AATCAGTTTG 1980  
 GTATGAATTT ATATCAATCA TATGATGTTT ATCCTAGTTT AGATAATAAG AGAGTAGCAT 2040  
 30 TTGATCTTCG TAGCAAAGAG GCCTTTCACG GTGGCCACAC AAAAGGAACA ATCAATATAC 2100  
 CATACAACAA AAACTTTATT AATCAAATTG GTTGGTACTT AGATTTTGAA AAAGATATAG 2160  
 ATGTAATTGG AGATAAATCT ACTGTTGAGA AAGCGAAACA CACTTTACAA TTAATTGGGT 2220  
 35 TTGATAAGGT AGCAGGCTAT CGTTTGCCAA AATCAGGCAT TTCAACCCAG TCCGTTTATA 2280  
 GCGCTGATAT GACAGGTAAA GAAGAACATG TATTAGACGT ACGTAATGAT GAAGAGTGGG 2340  
 40 ATAATGGACA CTTAGATCAA GCAGTTAATA TTCCGCATGG TAAATTATTA AATGAAAATA 2400  
 TTCCTTTTAA TAAAGAGGAT AAAATATATG TACATTGTCA GTCAGGTGTT AGAAGTTCAA 2460  
 TTGCAGTGGG TATATTGGAA AGCAAAGGTT TTGAAAATGT GGTGAATATT AGAGAAGGCT 2520  
 45 ATCAAGATTT TCCAGAATCA TTAAAATAAT TTAAGGATGT GGAAAAATG AATAAGCATT 2580  
 ATCAAATTGT TATTATTGGT GGCGGTACAG CAGGTGTTAC CGTAGCATCA AGACTATTAA 2640  
 GAAAAAATCA AAACCTAAAA GAGAAAATAG CAATTATAGA TCCAGCAGAC CATCATTACT 2700  
 50 ATCAACCATT ATGGACGTTG GTTGGTGCAG GGGTATCTAG TTTGAAAAGT TCTCGTAAAG 2760  
 ATATGGAAAG TGTTATACCT GAAGGTGCTA ACTGGATAAA ACAGGCTGTT TCAAGTTTTT 2820

55

TAGTAGCTCC AGGATTACAG ATTAATTGGT CTTCAATTAA AGGACTAAAA GAAAATATAG 2940  
 GTAAAAATGG TGTTCGTCT AACTATTAC CTGACTATGT TAACgAAACT TGGAAACAAA 3000  
 5 TTTCTAATTT TAAACAAGGA AATGCCATTT TTACGCATCC AAACACTCCT ATAAAGTGTG 3060  
 GAGGTGCgCC TATGAAAATT ATGTATTAG CTGAAGATTA TTTTAGGAAA CATAAAATCC 3120  
 GT 3122

10

(2) INFORMATION FOR SEQ ID NO: 520:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3982 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 520:

ATANAGATAT AnATnAATAT ATTGAGGTCA AACGATGATA ATTAAAATTT TAACAATTCT 60  
 ATTACTACTT TGTATATTGA GCTATTTGGT TACAAATAGA AAGAAGCCTT TTCTGTTCTT 120  
 25 AAAGACACTC TTTATGGGTG TGGTATTTAT CTTTATAGGA TATATTTCAC TGGCAATATC 180  
 TGCCGTAATT ATTTATGGTA TTATTCAATT TATCACAATT GATTTTGGTA GTTTTTTCTT 240  
 AATGGGTATT ATATTGATCT TGATTTCAAG TATATTCCAA TTATTTATAG TTAGATTACT 300  
 30 TTTTAGAAAA AAGAATGTCG ATTTGACAGA GGTTGTCGTT TTAsAsCATT TAATTCAATG 360  
 GTTCTTAGTT TACTTTGCGA TCTATCAAGC AGTAAATGAA AAAATGGACA TTAATGATAT 420  
 TAATATCGAC AATTTCCAAT CTGTCTTTTT TGACGTGTCT AATTTGAATT TAGTAATTCT 480  
 35 ACCAACGTTA ATCATTAGCT GGGTCACAAT ATTTAACTAT AGAATGAGAA GTTACAAATA 540  
 AAATCTATGA GATTATACCT TCAGACACCA ACATTCAAAT GGTGTCTTTT TTGTTGTGTG 600  
 40 GTTTIATTTT TGAAATTGCA AAAAGTAGAG GCATGAATTT TTTGACTAGT GTATAAGTGC 660  
 TGATGAGTCA CAAGATAGAT AGCTATATTT TGTCTATATT ATAAAGTGTT TATAGTTAAT 720  
 TAATAATTAG TTAATTTCAA AAGTTGTATA AATAGGATAA CTTAATAAAT GTAAGATAAT 780  
 45 AATTTGGAGG ATAATTAACA TGAAAAATAA ATTGATAGCA AAATCTTTAT TAACATTAGC 840  
 GGCAATAGGT ATTACTACAA CTACAATTGC GTCAACAGCA GATGCGAGCG AAGGATACGG 900  
 TCCAAGAGAA AAGAAACCAG TGAGTATTAA TCACAATATC GTAGAGTACA ATGATGGTAC 960  
 50 TTTTAAATAT CAATCTAGAC CAAAATTTAA CTCAACACCT AAATATATTA AATTCAAACA 1020  
 TGACTATAAT ATTTTAGAAT TTAACGATGG TACATTCGAA TATGGTGCAC GTCCACAATT 1080

55

## EP 0 786 519 A2

	TCAAAATCTT	GTGAGAGAAT	TTGAAAAAAC	ACATACTGTC	AGTGCACACA	GAAAAGCACA	1200
	AAAGGCAGTC	AACTTAGTTT	CGTTTGAATA	CAAAGTGAAG	AAAATGGTCT	TACAAGAGCG	1260
5	AATTGATAAT	GTATTAAAAC	AAGGATTAGT	TAAATAAAAC	TTCAATCGTT	GCTGTTATCT	1320
	GGAAATAATT	AATTAAATGT	TATGTTAATT	TTTGTTAATG	AAAAAAGTAA	TCTATTTAAT	1380
	GACAGGTAA	TGTAATTGTC	CTGAAATTGA	CTATATACTC	AGTAAGTATC	AATTTTAAGG	1440
10	AGAGCTTATA	ATGAAATTTA	AAAAATATAT	ATTAACAGGa	ACATTAGCAT	TACTTTTATC	1500
	ATCAACTGGG	ATAGCAACTA	TAGAAGGGAA	TAAAGCAGAT	GCAAGTAGTC	TGGACAAATA	1560
15	TTTAACTGAa	aGTCAGTTTC	ATGATAAACG	CATAGCAGAA	GAATTAAGAA	CTTTACTTAA	1620
	CAAATCGAAT	GTATATGCAT	TAGTCGCAGG	AAGCTTAAAT	CCATATTATA	AACGTACGAT	1680
	TATGATGAAT	GAATATAGAG	CTAAAGCGGC	ACTTAAGAAA	AATGATTTTCG	TATCAATGGC	1740
20	TGATGCTAAA	GTTGCATTAG	AAAAAATATA	CAAAGAAATT	GATGAAATTA	TAAATAGATA	1800
	ATAAATAAAA	CAGGTTGAGA	CAAAAAATGG	TCTTAACCTG	TTTTCAATTT	GCATATGTGA	1860
	TAAATTCTAT	ATCAAAATGC	TTATGTATAA	TGAATGACAT	TTAAAAGTAG	GGGAGACAAA	1920
25	TATAAATACA	ATAGTTCCTA	GGATTACTCT	CAAAATAACT	ATATCAATTA	TTTACTTTGC	1980
	TCTCCTATTT	TTTAAATAT	GTACATGTTT	AAACAATCAA	AAGTGTACAA	TATTAAATTA	2040
	TCATTTCCAG	TTCTAGTGCT	ATATTGGTAG	TAGTTGACTA	AATGAAAATA	AGCTTATAAC	2100
30	AAGTTTTTTC	AATACTCGTG	GGGCCACAAC	AGAGAGAAAT	AGGATCACCA	ATTCCAACAG	2160
	ACAATGCAAG	TTGGCGGGGC	CCCAACATAG	AGAAATTGGA	TCACCAATTT	CAACAGACAA	2220
	TGCAAGTTGG	GGTGGGGGCC	AACACAGAAG	CTGGCGAAAA	GTCAGCATAC	AAAAATGTGC	2280
35	AAGTTGGCGG	GGCCACAACA	GAGAGAAATA	GGATCACCAA	TTCCAACAAA	CAATGCAAGT	2340
	TGGCGGGGGC	CCCAACATAG	AAGCTGGCGA	AAAGTCAGCT	TACAAAAATG	TGCAAGTTGG	2400
40	CGGGGCCCCA	ACATAGAGAA	ATTGGAACCC	CAATTTCTAC	AGACAATGCA	AGTTGGGGTA	2460
	GGACATCGAT	AAAGAAATAC	TTTTTCTTTA	GCAATTAGTA	TTTCTTATGC	ATGAGCTTTA	2520
	CTCATGTATT	CATTTTTTTAA	GTACaCATTa	GCTACAGCTA	ATGATAAAGA	ACCACTACAT	2580
45	AATAAATCAT	TAGTGGTTCT	TTATCATTTT	TATCTCACTC	TTTTACTGGA	AGAAAAAGTT	2640
	TACGTTTGTA	GAACATGCCA	CAATACCAAA	AATAATTAAG	AAAAATAAGA	CGATAAGCAT	2700
	GATGACACTT	TTCAAACAAC	CTCTATCAGT	TTCTTTCGAT	TTTCTTTGTT	GAACCTTTTT	2760
50	ATAATCTTCA	AGTAGTTTTG	CGGCTTTTTT	ATTTATATGT	TTATTCATGA	TGTTGACTCC	2820
	TTATAATATA	TGTTTAATTC	ATTAAAATAG	TTGAAAACAT	GACTTGAAAT	AAAGATATAA	2880
55							



EP 0 786 519 A2

	AATTTGATGA TGATATTTGC TTTTATTTT CCAAATGGAA TTTACTTAAA CTGATGCATT	3000
	AAAATATTAA TGAAGCACTA GAATACATAA ATGAATAGTa ATGGTGcACA GTATAGAATA	3060
5	ATTAAGGCTA TATTAAGTAT AAATATCGTT AACTGTAAGC TATCTTTAGT TTTAATATAA	3120
	ACTATTAGGA TAATCGACGT AAGAAGAATC ATATATATTA ATGATGAAGA AGTCCATACA	3180
	AAATCCGCAT CATTGTGTGT TAATAATGGG ACTATAATTA ATCCGAAATT AATCATGCAT	3240
10	GCTATATATA CTATAATGTT ATACACAATG TTAATTTTTG TTCACCACCT TATACTTCTA	3300
	TTTTAAAAAC TTCTTTATAA TGATATATTG TTTAATGTTG AAATAATTAG ATTATCTAAT	3360
	TTTCATTGTC TTTACATGTA AAAGGCTATA TATAGTATGC TCTTTATGAT TCTAAATGCT	3420
15	TTTTAATATT TAATGCTCAT CAACATTTGG ATTTTGAATA TTCAATTCAA AAACTTTATT	3480
	AGCTACGTCA ATTGTAAAAT CAGAACCATA GTTGACATGA GCTACTTTTA ATTTTCCATC	3540
	TAAATAATAG ATTGCGATTG CAACATCGTA AAATTCGTCA ATGACAAATA AACTCTTTTC	3600
20	GTTTGTTACA ACCTCATGCT CTCCTGAGTA TACAACGTTA ATTTCCCAAT CATTAAAAAC	3660
	CATTTGTTAA CCTCCTTGAA CATTTAAATT GATTCAACTT AAGTTTAACT TATTCATACA	3720
25	ACTTCGTACA ATATCTAGAT GAACATTAAT TGTATTTCTA GAAATCTTTT TCAATTATAT	3780
	GTAATAATTA TACTTTTAAA TTTCTTATTT CAGTATAGTT TTA AACGAT TTTAAAATAA	3840
	TTCTGCAAAT ATATTAACAC ATAATGTGTT CAAAAAGTTT TGAACAATTT CAAAACTTT	3900
30	ATATAAAGGG nTTGACAACA TGGATTCAAA TnTCTTATTT TAAAAATTAC CTCATATAGT	3960
	GTCATGTTAG CCAATTTTTA AG	3982

(2) INFORMATION FOR SEQ ID NO: 521:

35	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1353 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
40	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 521:

45	AGCTTGgaTG ATTTAAaTG GtCCtTTCCC AaCcTTAGAT AATGAAAGAT tTGATAATCC	60
	TGAATATAAA GAAGCTATGA AAAAATATcA ACAGAGATTT ATGGCTGmAG ATGAGGCTTT	120
	GAAGAAATTT TTTAGTGAAG AGAAAAAAT aAAAAATGGA AATACTGATA ATTTAGATTA	180
50	TCTAGGATTA TCTCATGAAA GATATGAAAG TGTATTTAAT ACTTTGAAAA AACAAAGTGA	240
	GGAGTTCTTA AAAGAAATTG AAGATATAAA AAAAGATAAC CCTGAATTGA AAGACTTTAA	300

55

GTTAGGTAAA ACATTTTATC AAAACTATAG AGATGATGTT GAAAGTTTAT ATAGTAAGTT 420  
 AGATTTAATT ATGGGATATA AAGATGAAGA AAGAGCAAAT AAAAAAGCAG TTAACAAAAG 480  
 5 GATGTTAGAA AATAAAAAAG AAGACTTAGA AACCATAATT GATGAATTTT TTAGTGATAT 540  
 AGATAAAACA AGACCTAATA ATATTCCTGT TTTAGAAGAT GAAAAACAAG AAGAGAAAAA 600  
 TCATAAAAAT ATGGCTCAAT TAAAATCTGA CACTGAAGCA GCAAAAAGTG ATGAATCAAA 660  
 10 AAGAAGCAAG AGAAGTAAA GAAGTTTAAA TACTCAAAAT CACAAACCTG CATCTCAAGA 720  
 AGTTTCTGAA CAACAAAAAG CTGAATATGA TAAAAGAGCA GAAGAAAGAA AAGCGAGATT 780  
 TTTGGATAAT CAAAAAATTA AGAAAACACC TGTAGTGTC A TTAGAATATG ATTTTGAGCA 840  
 15 TAAACAACGT ATTGACAACG AAAACGACAA GAACTTGTG GTTCTGCAC CAACAAAGAA 900  
 ACCAACATCA CCGACTACAT ATACTGAAAC AACGACACAG GTACCAATGC CTACAGTTGA 960  
 20 GCGTCAAAC CAGCAACAAA TTATTTATAA TGCACCAAAA CAATTGGCTG GATTAAATGG 1020  
 TGAAAGTCAT GATTTACAA CAACGCATCA ATCACCACAA ACTTCAAATC ACACGCATAA 1080  
 TAATGTTGTT GAATTTGAAG AAACGTCTGC TTTACCTGGT AGAAAATCAG GATCACTGGT 1140  
 25 TGTGATAAGT CAAATTGATT CTTCTCATCT AACTGAACGT GAGAAGCGTG TAATTAAGCG 1200  
 TGAACACGTT AGAGAAGCTC AAAAGTTAGT TGATAATTAT AAAGATACAC ATAGTTATAA 1260  
 AGACCGAATA AATGCACAAC AAAAAGTAAA TACTTTAAGT GAAGGTCATC AAAAACGTTT 1320  
 30 TAATAAACAA ATCAATAAAG TATATAATGG GCA 1353

(2) INFORMATION FOR SEQ ID NO: 522:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1987 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 522:

GTCTGAGTCA GGTGCTGTTT GGTTAGATGC TGAAAAACA AGTCCTTATG AATTTTATCA 60  
 45 ATTCTGGATT AATCAATCAG ACGAAGATGT AATTAAATC TTAATAACT TTAATTTCTT 120  
 AGGAAAAGAA GAAATTGATC GCTTAGAACA ATCTAAAAAT GAAGCACCGC ATTTACGTGA 180  
 AGCTCAAAAA ACATTAGCTG AAGAAGTAAC TAAATTTATT CATGGTGAAG ATGCATTAAA 240  
 50 TGATGCAATC CGTATTTTAC AAGCATTATT TAGTGGTGAT TTAAATCAT TATCAGCGAA 300  
 AGAATTAAAA GATGGATTTA AAGATGTGCC TCAAGTGACA TTATCAAATG ACACAACAAA 360

55

EP 0 786 519 A2

	TGTTAACAAT GGTGCGATTT ATATTAATGG TGAGAGACAA CAAGATGTTA ATTATGCTTT	480
	AGCACCAGAA GATAAAATTG ATGGCGAATT TACGATTATT CGTCGCGGTA AGAAAAAATA	540
5	CTTCATGGTT AACTATCAAT AAATATAATT GCATAGCTAA ATAAATTAGA GCCTACTCAT	600
	ATTCATTCCT AAGAATGTAA TGAGTAGGCT CTTAATGTAC TTTTCTGTCT GTAAATTATC	660
	TAAAGAAATC TCTATCGCCT TGGCCAGGTG ATTGACGTTT TGATCGGCTT TGACGTTTGT	720
10	GTTGTTCTTT TTGTTGTTTT AATTTCACCTT TAACTTCTTT TGTTTTACCA TCACGGATAA	780
	CGGTAACAGT GACTGATTCA CCAGGTTTTT TATTTTCATA TAAATAGCTT CTTAAATCAA	840
	CATCATCTTT AATTTTCTTG CCATCAATTT CTGTAATAAT ATCACCTTTT TTAAGATCAA	900
15	TATCACTATC AGCTTTGGCG ACATAAATAC CGTCTTCTCT ATCAGTATGA AGTTGCTCGC	960
	GCTCTTCTTC AGGAATATCT TTCAAATTAA TTAAACCAAT ACCAATCGAA GGGCGGTCAA	1020
20	TTTTACCATG TTTTACAAGT TGTTCAATTG TTACTTTAAC TTCATTACTT GGAATAGCAA	1080
	ACCCGATACC TTCAACTTGT GTCGCAGCAA TTTTCATTGA GTTAATACCA ACTAAATTAC	1140
	CATTAATATC TACTAATGCG CCACCTGAGT TACCTGGGTT AATAGCAGCA TCTGTTTGAA	1200
25	GAACGCTAAC TTTTGTATTG CCACCAGTTG TCTCAGCGTC AATCGTACGT TCGCTTGCTG	1260
	AAATGATACC AGATGTTACA GAGTTAGCAA ATTGTAATCC TAATGGGTTA CCCATTGCGA	1320
	ATACGCTATC GCCAGTTTGT ACTTTTGAAG AGTTGGCAA TTGAATCGCT TTAATACCTT	1380
30	TTGTATTTTC AATTTTAAAGT ACAGCAATAT CAGTTACTGc ATCTTTACCA ACTAATTTG	1440
	CTTTAACTTG TTTTTTATTA TGTAATTGGa CTCTAATTTT ATTTGCGCCA TCAATAACAT	1500
	GATTGTTTGT AACGATATAA GCTGAATTGT TGTTTATTTG aTAGATAACA CCTGAACCTA	1560
35	CTCCAGCTTC AGATGGTTTA GATGATTTGC CyTTTAATAA GTCGTCTACA CTTGATGCTT	1620
	TTtGCATGtK AATAACTCCA ACAATTGTAG GGGAGACAGA TTTTATCATT TCATGAACGG	1680
	TACCGAATTT CTTGCTTTGA CCGTCTAATT GATTGCCACC TTTATTATTT GTTGTCTGAA	1740
40	CAGTTGAACC ATCTTTATTT AAAATTGTAC TATTTAATAC TTGCGCTATA CCAAGTACTA	1800
	GAAGTGCACC AATAATTCCA GCAATCAATG CAACGATGAC TGThTTAAAC CATGGAAATT	1860
45	TAGGTCTTCT GTATCTAGGT GTTTGGCTAT GGTThGTTGT AGAATGATCT GTATGATTAA	1920
	AATCTGACAT ACTTAACCTC CATTATATGA TTTATATATG CTTTAATTAT GTCTTTThTT	1980
	TATAATT	1987

50 (2) INFORMATION FOR SEQ ID NO: 523:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1351 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 523:

	TTAGAAGTCA AATCATTAnT GGCGTCnTAT CGAGTATACT ATTAAC TTCA ACTATTTT TAG	60
10	CAATTGCATA TATTTTAA TG GTTTAA CG GCCATATGAC aCTAACTTTG ACCTTAACGA	120
	CAATAATTAC AAGCTGTTTA ACCTTATTAA TATGTAGTAT TTTTATTAAT CCAC TTATAC	180
	AAAAAATTAA GCAGTTTAAT ATAAAACTA AGCAATTTGC TAACGGAAAT TACGCAAGCA	240
15	ATGATAAAAC GTTTAATTCA CCAAAAGAAA TTTATGAATT AAATCAATCT TTTAATAAAA	300
	TGGCTTCTGA AATTACGCAA CAAATGAATC AAATTAAATC CGAACAACAA GAAAAACAG	360
	AACTGATTCA AAAC TTAGCC CATGATT TAA AACACCTTT AGCAAGCATT ATTT CATATT	420
20	CTGAAGGACT ACGTGATGGT ATAATCACTA AGGATCATGA GATTAAAGAG TCATACGACA	480
	TATTAATTAA ACAAGCAAAC AGATTATCAA CATTATTTGA TGATATGACT CATATTATCA	540
	CTTTAAATAC AGGTAAACA TATCCCCCAG AATTAATACA ACTAGACCAA TTACTTGTAT	600
25	CAATATTGCA ACCATATGAG CAACGTATCA AACATGAAA CCGCACATTA GAAGTGAATT	660
	TCTGTAACGA AATTGATGCA TTTTATCAAT ATCGAACGCC ACTTGAGCGT ATTTTAACAA	720
	ACTTACTTGA TAATGCGCTA AAATTTTCAA ATGTTGGTAG TCGCATTGAT ATTAATATTA	780
30	GTGAAAACGA AGATCAAGAT ACTATCGACA TTGCTATTAG CGATGAAGGT ATTGGCATT A	840
	TACCAGAACT ACAAGAACGT ATATTCGAAC GTACATTCAG AGTAGAAAAC TCTCGTAATA	900
35	CAAAAACGGG TGGT TCTGGA TTAGGCTTAT ATATAGCTAA TGAAC TCGCG CAACAAAATA	960
	ACGCAAAAAT CAGTGTAAGC AGTGATATAG ATGTAGGAAC TACGATGACT GTAACATTAC	1020
	ACAAATTAGA CATTACGTCA TAATCCGATT TATTTATAAA ATAAATGCA AAGACTAAAA	1080
40	AGAAGCTCCC ATTAATGAGG GCTTCTTTTT TTGTTTATTT AGAATAAACT TTATGGGTAT	1140
	CCTTCTCATC ATTTTCAAGA CTTGAAAGAT TTGTAGCTTG AATAATATAT TTAGGACGTG	1200
	CCTTAACTTC ATAATATATC CTGCCAATAT ATTCACCTAC AACACCAATT GAAATTAAC T	1260
45	GTATGCCGCC TAATAATAAA ATAGCTGCAA TCGTTGAAAA ATATCCCGGA ATATTAACAC	1320
	CAGATATCAT AATATTGATG GAATAAATAG A	1351

(2) INFORMATION FOR SEQ ID NO: 524:

50

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 433 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 524:

5 TTGTTGTCAG TTTAACGCAA CGTGTTACCG ACTTTTCAAG TACACAACAA TATATATGTC 60  
 TATTCCATT TTAGCCCCcTG CCATTTTTCAT CATTGGTGGT ATTATGTTGT TTATTTCAAC 120  
 ATTTAATAGC TTAGATGAAA CTGCTGAAAA TAATAACAAA ATAAAGAAAC TAATGATTAA 180  
 10 AGGACTTATC ATTATTAACA TTTCATTTAT CGTTATGATG GTTTTAACAC CATATTGGTA 240  
 CTGTATTFTA ATCGTCTATC TTATTTTCTT GTTGTTTTFA TTGTGGCAAA AGGTTTATAA 300  
 ATTTTAATAC CAAACTATT AAACACTTCT GATATTCTTA GTTCAAAATA TCAGAAAGTG 360  
 15 TTTATAGTGT TATCTAGTTC AGATAAATAT TTCCTTACTT AAAAAACGC CCTCCTCTTA 420  
 TTTTGACCCC nAT 433

## (2) INFORMATION FOR SEQ ID NO: 525:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1845 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 525:

30 CCCCCACTAT GATATGCTTG GCCTATTGCG TCAATCCCCT TATATTTCGG CAATAGAGAT 60  
 GGATGTATAT TCAATATTTT ACCTTCAAAT GAAGCTAATA AGTCTGGACC AATTAGACGC 120  
 ATGTAGCCAG CTAGAATAAT CCACTCTACC TTATCTTCAT TTAATAATGT TACTAAATGT 180  
 35 TGTTCATACG CTGCTTTTGA ATCAAATTGT TTTGGTTCAT TAATATAAAC AGGAATATCG 240  
 TGCTTTTTTG CTCTATCTAT ACAAACGCA TTTTGATGAT CCGTATATAG CGCCGTAAC 300  
 TCAATATTTT CAAGTTTTCC TGATTCAACA TGCTCAACTA TATTTKCAA GTTACTTCCT 360  
 40 GAACCTGATG CAAAATCGC AATTTTAACC ATTGTTATAC CCCCACAAT TCAATTGCAG 420  
 TTGACTCATT TTTCACAATA TGACCAATTT GATAGGCTTC CACATTTGT TCTGCTAAAA 480  
 TCTCAAAGC GCGTGATACA TCTTTTTCAT CAACGATAAC CGTATAGCCA ATACCCATAT 540  
 45 TAAAAATGTT ATACATTTCA TTTGTGTCTA TATTGCCTTG TTGTTGTAAC CAATCAAATA 600  
 TTTTGGTGT TGGAAATGAT GTAGTATCAA TTCTAGCAGC ATATCCGGCT GGCAATGCAC 660  
 GTGGAATATT TTCATAAAAA CCTCCACCAG TAATATGATT CATTGCCTTA ATAGAAACTT 720  
 50 CTTTTTTCAA AGCAAGTACA GGTTCGACAT ATAATTTAGT TGGTTCTAAA AAGACATCGA 780

55

5 GCACTAAACT GTATCCATTT GAATGAATGC CACTTGACGC AAGCCCTATA ACAACTTGTC 900  
 CCTCTTTCAC TTCTGAACCA TCTACATAGT CATCCTTTTC AACTGCTCCA ACAGCAAATC 960  
 10 CAGCTACATC ATATTCGCCT TCGTGATACA TTTCACCCAT TTCAGCAGTC TCTCCACCGA 1020  
 TAAGTGCAGT ATTCGTTTCA ACACATGCAT CACTAATACC TTAAACAATT TGTTCATAAA 1080  
 CTTCAGGAAC AACTTTGTTT GTAGCAATGT AATCTAAAAA ATATAATGGT TCTGCACCTG 1140  
 15 TCGTTAAAT ATCATTAAAC CACATTGCGA CTGCATCGAT ACCTATCGAA TCATGTTTAC 1200  
 CATAGTCGAT AGCTAGTTTT AATTTCGTAC CTACTCCGTC TGTTCAGAA ACTAAAACTG 1260  
 GCGCTGtCAT ATTTAATTGT GATAAATCAA ATGTAGCACC GAATCCACCT AAACCACCGA 1320  
 TAACTTCTTT ACGCATCGTA CGtTTAAcaw GaCtAGaCaT TctTTtACA GCTTCATAAC 1380  
 CAGCATGAAT ATTTACACCA GATTGTTTCAT ATGCTTTAGA CATTTAAATT TCCCTCGCTA 1440  
 20 TCAAAAAAGT GTTTGTTTTT AGAAATATAT TGCTTTTGTC GATGACTTAA ATGCGCTTTA 1500  
 TAATTTGCTT CATAATCATA TAACCCTGCA GGATAATCTC CAGTGAAACT TTCTACACAT 1560  
 AAGCCACTAT ATGGCGCGTC ATAATCTAAA CCAATTGATT CAATTAACCC ATCTACAGAT 1620  
 25 AGATATGCTA ATGAATCAGC GCCAATATAA TCTTTAATTT CTTCAGGTGA TTTGCTTGCA 1680  
 GAAATTAATT CTGCCGTAGT TGAAACATCG ATTCCGTAAA AACTTGAAA CATAAATTC 1740  
 GGTGATGCTA TACGCACATG CACTTtATTT GCACCAGAAT CkTTTAACAT TTTACAATG 1800  
 30 CGTCAATTG TCGTACCGCG nACAATGGAA TCATCAACAA GACTG 1845

## (2) INFORMATION FOR SEQ ID NO: 526:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1141 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 526:

TCATAAAATnG TAAATGTTAA TCGTCATATA ATATTAAATTT ACAACACCAT TTTGGTTATT 60  
 TGAAGCTTGT GCGCTTGTT GTGTGCCACC TTGATTTTGA TTTGAGTTT GATCTGTAGC 120  
 45 AGGTTGTTGT TGATTGCTGG AmtCACTGTT ATTAGTTGAm TCACTGTTCT CGTTAGATGT 180  
 CTTATCTTTA TCTGTCTGAT CATTTTGTTC TTTTCTCAAT AACTATTAT CTAAAGGCGT 240  
 50 TAATGGTATT AATGAACCAT AATGATTAAAT GACACGTTGA TCTAAGAAAT CATTTTATC 300  
 ATTAATAGGT GATAATTCTA AGTCTTTACG AAGTAAGTTT GCATATTTT GAATGCTTTC 360

55

AGTTTAAATT TCAACATCAT TTGTTaAGTA CATTTTGTCT AAAGCTTTAA TTTCAGAGTT 480  
 AGTTAAATTA TGCTTTGCAT TTTTACCTAC AATTGAATC ACGTTATCAA GTTTATCAAT 540  
 5 AGAATCAACT TCCTGTGCTT TTTGGAATAA AATCTTAATT AATCCATTT GACGTTGTCC 600  
 ACGTTTAAAG TCTGAATCAT GATGTCTAGT TCTAGCAACT GCTAAAGCCT CATCACCATT 660  
 10 TAATTTTGG TACCCTTTT TAATTTAAT CTTACCAGTA TCATCTGTGT TAGGTTTATT 720  
 TAAGTCGTAT GGCACATCAT AGTATATGCC ACCAAGCTCA TTTACAGCCT CGACAAATGC 780  
 TTTTATATTG ACTCTCACAT AATAATCAAC AGGTACATTC ATGGTAGCTT CTACCGAATC 840  
 15 CATTGCGGCA ATTGGACCAC CATATGCATG TGCATGGGTA ATCTTATCGT AATAGCCAAC 900  
 TTTAGGAATG TAGCTGATAG TATCACGTGG AATACTAAGC ATTCTAATTT GATGTTTTGA 960  
 TTGATTAAAA GTAGTTAAAA TCATAGCGTC TGATCTAGAG TGTTCAGCAT CCTGTCTTTT 1020  
 20 TTTTCTTCTT CCATCGTTAT CATCGATACC TAAGAAAAGA ATAGAGATAG GTTGTCTTTC 1080  
 GGGATGACTT TATTATCTCT TAAGTTGGAT TGACGTTAGC ATTTTGTCTG TCTTGAGAAG 1140  
 A 1141

25 (2) INFORMATION FOR SEQ ID NO: 527:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1565 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 527:

GCACCATTAG CATTGGGcnc ACTGATTGGA GTAnCAGTTG TTGAAAATTC GGCGCCAACA 60  
 AGTAAACmGG CmCaGGCaGC mATamCCCCa TATTATAcTT ATAATGGTTA TATTGGTAAT 120  
 40 AATGCTAATT TTATTCTGGA TAAGAATTTT ATTAACGCGA TTAAGTATGA TAATGTGAAA 180  
 TTCAATGGTA TTAAATTAGC TAAAACGAAT ACGATAAAAA AAGTAGAGAA ATATGATCAA 240  
 ACTTTTAAAG GTGTTTCTGC AAAAGGAAAC GAAGCAAGTC AATTGCAATT TGTAGTTAAA 300  
 45 AATAATATTT CATTAAAAGA TATCCAAAAA GCTTATGGCA AGGACTTGAA AAAAGAAAAT 360  
 GGTAAAACAA AGGAAGCTGA TAGCGGTATT TTTTACTATC AAAATGCTAA AAAGACATTA 420  
 GGCATCTGGT TTGTCGTTGA TCATAATAGA GTTGTCTGAG TAACAGTTGG ACATACACCA 480  
 50 TACAAAACAA GTAAATAAAA TAATGGCATA TTAAGGCTAG AGTGTGAGGA GTGATACCGC 540  
 ATTCTAGTCT TTTTIATTAA ATAATAACGA TTATTGCGTC TTACATAGTT GTTTGAAATT 600

55

	GATTAAGTAT ATAGAGCACT ATTTTGTATT TGTTAATATT TTCACAAAA TAAAGCCTTG	720
	ATAAATTTTA AAATATAATT AAGCTCAATT TTTAAATTT TATTTAGCTA CAGATAACAT	780
5	TTTTAAAAA GAAAAGAATC AATAAATAAA ATCAACGAAC AAAAAGTATA GAAATAAATA	840
	GAAATAATCA TTTACTTTTC TGAAAAATTA AATTAATATT TTATTTATAA GCTGTTTTTA	900
	AGATTTTCAGG AGGAATGAAA TGTGaGGAAA TTTTCAAGAT ATGCATTAC AAGTATGGCA	960
10	GCATTAACCT TGTTGAGCAC TTTATCACCA GCAGCATTAG CGATTGATTC AAAAAATAAA	1020
	CCAGCTAATT CTGATATTAA ATTTGAGGTG ACTCAAAGA GTGATGCGGT CAAAGCATT	1080
15	AAAGAATTGC CTAAATCCGA AAATGTAAAA AATATTTATC AAGATTACGC TGTTACTGAT	1140
	GTAAAACTG ATAAAAAAGG ATTTACGCAT TATACATTGC AACCGAGTGT TGATGGTGT	1200
	CATGCACCTG aCAAGAAGT GAAAGTACAC GCAGACAAAT CAGGAAAAGT CGTTTTAATC	1260
20	AATGGGGATA CTGATGCGAA GAAAGTAAAG CCAACGAATA AAGTGACATT AAGTAAAGAT	1320
	GACGCAGCCG ACAAAGCATT TAAAGCAGTT AAGATTGATA AGAATAAAGC GAAAAATCTT	1380
	aAAGATAAAG TCATTAAAGA rAACAAAGTT GAAATCGATG GTGACAGTAA TAAATACGTT	1440
25	TATAATGTTG AGTTAATTAC AGTGACACCA GAAATTTAC ATTGGAAAGT TAAATTGAT	1500
	GCTCAAACCTG GCGAAATTTT AGAAAAAATG AACTTAGTTA AAGAAGCTGC AGAACTGGT	1560
	AAAGG	1565

(2) INFORMATION FOR SEQ ID NO: 528:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2870 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 528:

	GGATAGGTAA ATCCATTCAA AATTGGTTGT ACTCATTTTA AATCATATTT TTTAATATAG	60
	AAACACATAA TTAAAGGAGT GATGATATTA ATGAATAAAC TTCGAGACAC TACTTTTCTA	120
45	TCATATTTAT TTAATATCAT ATTGTGGGGA TCTGCATTTT CAATGATAAA GATTGCGTTA	180
	AATGATTTTA GTGCAGAATC ATTGTCGGCA TTTCGTTTAA TTTTAGCAAC AATAATTCTC	240
	TTGCCGTTTC TAATTATAAA GAAATTGCCT ACCCCTGAAC TAAGAGATAT CCCTGTTATT	300
50	TTTATTTTAG GATTTTGTGG ATTTGTGATA TATCACACAG CTTTAAATTT TGGTGAAaCT	360
	TTGATTAGTG CAGGTATATc TGGrATTcyA GTCTCTACAA CGCCTATTTT TTCTAGTGCT	420



	GCATTTATAG GAATATCCAT TATTTCAATA AGTAAAGATG ATTACACAAC TATTAATGTA	540
	TTAGGTGTTT TTATTATTTT ACTTGCACTT TTTAGTGAAA GTTTGTATTT TACTTTCCAG	600
5	AAAAAATACA TAGAAAAATA CGGCTTCATC GCTTTCACAC TATaTACAAT AATGGCAAGC	660
	TCACCATTTA TGCTTATTTT TATTCCTGAA ATCATCAACG ATATACACGG CGCCACTTTT	720
10	ACATCAATAG TATCGGTACT TTATTTAGCT ATATTCCTTA CTATAATTCC ATACGTTTTG	780
	CTTGCTTATA TTGTGAAGTC AGTTGGTGTC TCTGATGCAA CAATGTCTCT TTATTTAACA	840
	CCTATCGTTT CTTTATTATT ATCTTATCTG TTATTAGACG AGCTACCAAC AACCCTTGCT	900
15	ATTATAGGCG GAATTATCAC TCTACTAGGC GTTAGTTTAA GTAACCTCTT TCAAAATACA	960
	TAATTATTCC AAGTCCCGCA CCTCAGAATC CAAAACATT CGAGTGATAA AATTTTAAAA	1020
	ATCAAAAATA TAAAAATGAT CTAATTTCTG CAAATTTACC AATATAAATA CTAATATTTG	1080
20	CAATTCACAA AGGGGTATAG TCTGAGTGTA TTCTAATACG AAAGGACTTG GTGGATATGT	1140
	ATTACAGTTA TGGAAATTAT GAAGCATTG CGCGCCCTAA AAAACCTGAA AATGTAGAAA	1200
	ACAAATCCGC TTACTTAATC GGATCTGGTC TAGCTTCACT TGCTGCAGCT TGTTTTTTAA	1260
25	TAAGGGATGG TCAAATGGAA GGTTCGAAGA TTCATATTTT AGAAGAGTTA CCTAAAGCAG	1320
	GTGGTAGTCT TGATGGTGAA AATATGCCTT TAAAAGGCTA TGTGTCCGC GGTGGTCGTG	1380
	AAATGGAGAA CCACTTTGAA TGTGTGTTGG ACTTATTCAG ATCTATCCCT TCATTAGAAA	1440
30	TCGATAACGC GTCTGTATTA GATGAGTTCT ATTGGCTAAA CAAAGAAGAC CCTAACTATT	1500
	CTCGCTGTCG TGTTATTGAG AAACAGGGTC AACGTTTAGT CACAGACGGA GACTTCACTT	1560
35	TGACTAAAAC GGCATTAAA GAAATTTTAG ATTTATGCTT AACGAATGAA GAAGATTAG	1620
	ATGATGTCAA AATAACAGAT GTATTTTCCG ATGACTTCTT TAATTCAAAC TTTTGGATTT	1680
	ACTGGAAAAC GATGTTTGCA TTTGAACCGT GGCATTCTGC AATGGAAATG CGTCGCTATC	1740
40	TAATGCGATT CGTTCATCAT ATTAGTGGTC TCGCAGACTT TTCAGCTTTA AAATTCACTA	1800
	AATATAATCA ATATGAATCT TTAGTATTAC CTATGGTTGA ATATTTAAAA TCGCATGGGG	1860
	TTCAATTTGA ATACGATGTA AAAGTCGAAG ATATTAAAAT AGATGTTACG ACAAGTCAAA	1920
45	AAATTGCCCG AGAAATATTA ATTGACCGTA ATGGTAATGC AGAATCTATT AAAGTACTA	1980
	TAAACGATCT TGTCTTTGTG ACAAACGGTA GTATTACAGA AAGCTCTACT TATGGTGATA	2040
50	ATGATACACC AGCGCCACCA ACTGACGAAT TAGGTGGTAG TTGGACACTA TGGAAAAATT	2100
	TAGCGCGACA AAGTCCTGAA TTTGGTAATC CTGATAAGTT TTGCCAAAAT ATTCCTAAAA	2160
	AAAGTTGGTT TGTTCAGCA ACTTCTACAA CAAACAATAA AGAGATTATC GATACAATAG	2220
55		

TCAATGATTG TGCATGGCAA ATGAGTTTTA CAATCAATCG TCAGCAACAG TTTAAAGACC 2340  
 AACCTGAAAA TGAAATATCT ACATGGATTT ATGCCTTATA TTCAGATGTA AACGGCGATT 2400  
 5 ATATTAATAA GCCAATTACA GAATGTAGTG GTAATGAAAT ATGCCAAGAA TGGCTGTATC 2460  
 ACTTAGGTGT ATCAACTGAC AAAATTGAAG ACTTAGCAAA ACATGCATCT AATACGATTC 2520  
 CTGTTTATAT GCCATATATC ACATCTTATT TCATGACGCG TGCTATCGGC GACAGACCTT 2580  
 10 TAGTCGTCCC GCATCAATCT CAGAACTTAG CATTTATTGG TAACTTTGCA GAAACAGAGC 2640  
 GAGACACTGT ATTTACAACA GAATATTCGG TTCGTACTGC CATGGAAGCT GTTTATCAAT 2700  
 TACTAAATAT AGATCGTGGT ATTCCAGAAG TCATCAATAG TCCATTGAT CTTCGCGTCT 2760  
 15 TAATGGATGC CATATACGAA CTGAATGACC ACCAAGATTT GCGTGAGATT ACTAAAGATT 2820  
 CGAAAATGCA AAAACTCGCA TTAGCAGGAT TCCTTAAAAA GATAAAAGGT 2870

20 (2) INFORMATION FOR SEQ ID NO: 529:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3105 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 529:

30 CnGTTTAGCT GCAAAAAGTA AATAACGACA CTGTATTTGG AATTTTGCAA TTAGAAACAC 60  
 TTTTGGGTGA CATTAACTCA ATTTTCAGCG AGATTGAAAG CGAATACAAA ATGTCTAGAG 120  
 AAGAAATTTT AATTTTACTA ACTTTATGGC AAAAAGGTTT TATGACGCTT AAAGAAATGG 180  
 35 ACAGATTTGT TGAAGTTAAA CCGTATAAGC GTACGAGAAC GTATAATAAT TTAGTTGAAT 240  
 TAGAATGGAT TTACAAAGAG CGTCCTGTTG ACGATGAAAG AACAGTTATT ATTCATTTCA 300  
 ATGAAAAGTT ACAACAAGAG AAAGTAGAGT TGTTGAATTT CATCAGTGAT GCGATTGCAA 360  
 GTAGAGCAAC AGCAATGCAA AATAGTTTAA ACGCAATTAT TGCTGTGTAA GTTTAATAGC 420  
 ATAAAAAGAG GTTTTCATTA AGTTGAAAAC CTCTTTTTGT TGTTGGCATT AATTTTTCAA 480  
 45 ATGTTGACTA CTCAATCCTA AATTATAAAT AGTATAGCGC aCAAaTGCTT AAGAAATTTT 540  
 TTCTATGGCA CAAATGAATG GAGCATGATT ACGTTGGTTT AAAAATTGAT ATTGCAAAAC 600  
 TTGCGCATGC TTTTGATCCA AAGTACTCAA GTAATCAAGC AATGCATGCT TCTCAATTTG 660  
 50 TCCTTCGCTA TGACCATGAT ATATAACAAG TACAATAATA CCTTCAATTG ACATTAATGA 720  
 TAGCAATGAA TTAATAGCTT GGATTGTCGT GTCAGGCTTT GTCACGATAG ATTTATCACC 780

55

	TATATGATGT TCAATATTTT CATGTCCATC TTTTATTAAA GAAACATGAT TGAAATCCTT	900
	AACTTTATCA CGTGTATTTT CCAAAGCTAA ATCTTGAATG TCGAAACCAT AAACATGTCC	960
5	TTCTGGTACT TGTTCCGGCTA AAAATAAAGT GTCATTGCCG TTACCGCAAG TTGCGTCTAC	1020
	AACAATACTT TCTGGTGTTA TATGTTGTTT AATAAGTGTT TTTGAAAAAG GGAGTATACG	1080
	TTCTAATTTT ATTGCTTCAC CTTAGACTTG TAACGCAAGC CTTGATAAGA ATTTCTACGT	1140
10	GCTAATTCAG CATCGATGCC ATTTAATACT TCCCATTTAT TAACACTCCA CATTGGACCT	1200
	ACCATGATAT CTATTGGACC ATCACC GGTA ATTCGGTGAA CGATCATTTT AGGGGGAATC	1260
	ACTTCTAATT GGTCCACAAC TAGGTTTGTG TACTCTTCTT GAGTCATAAA AGTTAATAAA	1320
15	CCTTTATCGT ATTGTTTTAC CATCGGTGTA CCTTTTAACA AATGAAGTAA ATGAATTTTA	1380
	ATACCTTGTA CATCCATTTG TGCacTTCTT TGGCAGTAGC CATCATCATG TCATAGTCTT	1440
20	CGCCAGGTAA GCCATTAATG ATGTGTGTAC ATACATTGAT ATTATGCTTA CGTAATTTTG	1500
	CCACACCATC ATAATAAGTT TTCATATCAT GGGCACGATT GATTAAATCA GATGTTGACT	1560
	GATGGATTGT TTGTAGTCCT AATTCAACCC ATAAGTATGT TCGTTGATTC AAATCTGCTA	1620
25	AATATTCGAC AACATCGTCT GGTAGACAGT CAGGACGCGT ACCAATAGAT AATCCCACAA	1680
	CACCCGGTTC TTTAAGTACA GGTTCGAATT TTTCTTTTAA TACTTCAACC GGTGCATGTG	1740
	TATTTGTAAA TGCCTGAAAA TAAGCAATAT ATTTTCCTTC GTGCCATTTT TCATGCATCT	1800
30	TTTCTTTAAT TTCTTTAAAT TGTACTGCGA TTGAATCTGC ACGATTACCT GCAAAGTCTC	1860
	CGCTACCTGC AGCAGAACAA AATGTACATC CACCATGTGC TACAGTGCCA TCGCGGTTAG	1920
	GACAGTCAAA CCCGCCATCC AATGCAACTT TAAATATTTT TTGTCCAAAT TTATTTTTTA	1980
35	AATGGTAATT CCATGTGTGA TAACGTTTGT TTTCAAAGC GTATTGGAAA TGATTGCCCA	2040
	TATGTCATTT TCCTTTCTAT AAAAAAGAG TTCTAAGTAC AGATTTTAAC ATATTTTAAT	2100
40	GTTATAGTGT TTATTATAGT TTGACAAAAA AGAGAGAGGA ACTATGAAAT ATGAATATAC	2160
	CTAAATCAGT CTGGTGGCTA GTAATTGGCA TGGCGTTAAA TATTACTGGT TCCAGTTTTT	2220
	TGTGGCCTTT AAATACAATT TATATGAAAC AAGAACTTGG AAAAAGTTTA ACTGTTGCTG	2280
45	GTTTAGTGCT AATGATAAAT TCATTTGGCA TGGTTATTGG AAAC TTATTA GGTGGTTCAC	2340
	TATTTGATAA ATTAGGTGGA TACAAGACGA TTTTAATTGG AAC TTCACT TGTCTTTGTA	2400
	GTACAACGCT ACTTAATTTT TTTCACGGGT GGCCTTGGTA TGCTGTATGG CTTGTAATGT	2460
50	TAGGGTTTGG TGGCGGAATG ATTATTCCTG CGATATACGC TATGGCTGGA GCAGTGTGGC	2520
	CAAATGGCGG AAGACAAACG TTTAATGCGA TATACTTAGC GCAAAATATT GGTGTGGCTG	2580
55		

ATCTTATTAT GTATGTTGTG TTTGCGCTTG TCGCGGTAAC GCAATTTAAT ATTGAAATTA 2700  
 ATGCGAAAGT TAAATATCCA ACTCATTTAG ATATTACTGG TAAAAAGAAT AAAGCAAGAT 2760  
 5 TTATTTTCATT AGTACTAATT TGTGCAATGT TTGCAATTTG TTGGGTTGCA TATATTCAAT 2820  
 GGGAGTCTAC AATCGCTTCA TTTACACAAT CTATTAATAT TTCAATGGCA CAATATAGTG 2880  
 TTTTATGGAC AATTAAACGGA ATAATGATTT TAGTAGCACA ACCATTAATT AAACCGATTC 2940  
 10 TCTATCTGTT AAAAGGAAAC TTAAAGAAGC AAATGTTTGT CGGCATCATC ATTTTATGT 3000  
 TGTCGTTCTT TGTACAGAGT TTTGCCGAAA ACTTTACAAT ATTTGyTGTC GGTATGATTA 3060  
 15 TTTTAACTTT TGGAGaATGT TTGTATGGCC AGCAGTTCCA ACTAT 3105

(2) INFORMATION FOR SEQ ID NO: 530:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 5532 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 530:

TAATGATTAA ACCAGATGCA GTACAAAGAA ATCTAATTGG TGAAGTAATT TCAAGAATTG 60  
 AAAGAAAAGG ACTAAACTT GTCGGTGGTA AATTAATGCA AGTACCAATG GAACTTGCTG 120  
 30 AAACACATTA TGGTGAACAC CAAGGTAAAC CATTTTATAA TGATTTAATT TCATTTATTA 180  
 CATCAGCACC AGTGTTTCGCA ATGGTAGTTG AAGGTGAAGA TGCAGTTAAT GTATCTAGAC 240  
 35 ATATTATTGG CAGCACCAAT CCTTCAGAAG CTTACCAGG ATCAATTAGA GGTGATTAG 300  
 GTTTAACTGT TGGTAGAAAT ATCAATTCACG GTTCAGATTC ATTAGAGTCT GCTGAACGTG 360  
 AAATTAATCT ATGGTTTAAT GAAAATGAAA TTACTAGCTA TGCTTCACCA CGTGATGCAT 420  
 40 GGTATATGA ATAAATATA AACTGTAAAC CTTTACGATT TATTTATAAA GG TAGAAAAG 480  
 GTTTTGTTAT GTGGTTAGTC ATTATGaTTA TACATAACAA GGCCCGTTTT TTATGTTGTA 540  
 GTAAATTACT TGAAAAATTT TATAGTTTTT kGGTAACACG TAtTaAAAAg AGAGGAATAT 600  
 45 TCTTTATCAA ATGAAACTAA ACAGAGAGAA GGGGTTGTTA AAATGAaGAA TATTATTTCC 660  
 ATTATTTkGG GGATTTTAAT GTTCTTAAAA TTAATGGAAT TACTATATGG TGCTATATTT 720  
 TTAGATAAAC CACTTAATCC TATAACAAAA ATTATTTTTTA TACTGACTCT CATTTATATT 780  
 50 TTTTATGTAT TAGTAAAGA ATTGATTATA TTTTGTaAGT CAAAGTATAA CAAAAGCGCT 840  
 TAACATATGT ATATTTTAAT ATCATAATTT TTTTAAACGG ACTGATTAAC TTTATTAATA 900

55

	GATACGATTA TATTAACG GCTAATCATT TTAAATTAAT GATTATATGA TGCAACTGTT	1020
	TAGAAATTCA TGATACTTTT CTACAGACGA ATATATTATA ATTAATTTTA GTTCGTTTAA	1080
5	TATTAAGATA ATTCTGACAT TTAATGAG ATGTCATCCA TTTCTTAAT TGAGCTTGAA	1140
	AACAAACATT TATGAATGCA CAATGAATAT GATAAGATTA ACAACATATT ATAATGTTAT	1200
	CGTGAAGTA TGAAAGGAGC GAGTGTGTAT GAGATACCTA ACATCAGGAG AATCACATGG	1260
10	ACCTCAATTA ACAGTTATTG TTGAAGGTGT ACCTGCAAAT ATAGAAATTA AGGTTGAGGA	1320
	TATTAATAAA GAAATGTTTA AGCGTCAAGG CGGTTACGGA CGTGGACGTC GTATGCAAAT	1380
15	TGAGAAAGAT ACAGTAGAAA TAGTATCAGG CGTTAGAAAT GGTATACAT TAGGTAGTCC	1440
	AATTACTATG GTTGTAACCA ATGATGACTT TACGCATTGG AGAAAAATTA TGGGAGCAGC	1500
	TCCAATAAGT GAAGAAGAAC GTGAAATAT GAAACGTAAT ATTACAAAAC CAAGACCTGG	1560
20	TCATGCAGAT TTGGTTGGAG GTATGAAATA TAATCATCGT GATTACGAA ATGTGCTAGA	1620
	GCGATCATCT GCTAGAGAAA CAGCAGCTCG AGTTGCAGTC GGTGCCTTAT GTAAAGTGTT	1680
	ATTACAACAG TTAGATATCG ATATATACAG TCGTGTGTGTT GAAATAGGTG GAATTAAAGA	1740
25	TAAAGATTTT TATGATTCAG AAACATTTAA AGCAAATCTT GATCGTAATG ATGTTCTGTG	1800
	AATTGATGAC AGTATCGCAC AAGCAATGCG AGATAAAAT GACGAAGCTA AAAATGAAGG	1860
	AGATTCAATT GCGGTGTGCG TTCAAGTTGT AGTTGAAAAT ATGCCTGTTG GTGTAGGTAG	1920
30	TTATGTGCAT TATGATCGTA AGTTAGATGG TAAGATTGCA CAAGGTGTTG TCAGCATAAA	1980
	TGCTTTTAAA GGTGTAAGCT TTGGTGAAGG ATTTAAAGCA GCTGAAAAGC CAGGTAGTGA	2040
35	GATTCAAGAT GAAATTCTAT ATAATAGTGA AATTGGTTAT TATCGTGGAT CTAATCACTT	2100
	AGGTGGTTTA GAAGGCGGTA TGTCAAATGG AATGCCAATT ATCGTTAATG GTGTAATGAA	2160
	ACCAATTCCA ACGTTATATA AACCATTAAA TTCAGTAGAC ATTAATACTA AAGAAGACTT	2220
40	TAAAGCAACA ATTGAACGTT CTGATAGTTG TGCTGTTTCT GCAGCAAGTA TCGTCTGCGA	2280
	ACATGTCGTA GCATTTGAAA TAGCAAAAGC ATTATTGGAA GAATTCCAAT CAAATCATAT	2340
	TGAGCAACTT AAACAACAAA TTATTGAGCG CAGACAATTA AATATTGAGT TTTAACAACA	2400
45	AGAACAATTG AGGTGTAATC ATGAAATTAC AAACAACATA CCCTTCAAAT AATTATCCAA	2460
	TATATGTTGA ACACGGTGCA ATTGACCATA TTAGCACGTA TATTGATCAG TTTGATCAAA	2520
	GTTTTATATT AATTGACGAG CATGTAAATC AATATTTTGC TGATAAATTT GATGATATTT	2580
50	TATCATATGA AAATGTACAT AAAGTTATTA TTCCAGCTGG TGAAAAGACG AAAACATTTG	2640
	AGCAATATCA AGAAACATTA GAGTATATTT TATCCCATCA TGTAACCTCGT AATACAGCAA	2700
55		

	ACGAGGCGTG CACTTTATAC AAGTGCCAAC GACTATACTA GCGCATGATT CTAGTGTGG	2820
	CGGTAAAGTG GGTATTAAC TAAAGCAAGG TAAAAACCTT ATCGGTGCAT TTTATCGTCC	2880
5	AACTGCTGTG ATTTATGATT TAGTCTTTTT AAAGACGTTA CCATTGAGC AAATATTAAG	2940
	TGGCTATGCA GAAGTTTATA AGCATGCGTT ATTGAATGGT GAATCAGCGA CGCAAGATAT	3000
	CGAACAGCAC TTAAAGATA GAGAGATATT ACAGTCATTA AATGGTATGG ATAAATATAT	3060
10	TGCTAAAGGT ATTGAAACGA AGCTGGATAT TGTTATTGCA GATGAAAAAG AACAAAGGTGT	3120
	ACGTAAATTT TTAATTTAG GTCATACATT TGGTCATGCT GTTGAATACT ATCATAAAAT	3180
15	ACCTCATGGT CATGCAGTGA TGGTTGGCAT TATCTATCAA TTTATAGTTG CGAATGCTTT	3240
	GTTTGATTCT AAGCATGATA TTAATCATT TATTCAATAT TTAATACAAC TCGGCTATCC	3300
	TTTAGACATG ATAACGACT TGGATTTTGA AACGTTATAC CAATATATGC TAAGTGATAA	3360
20	AAAGAATGAT AAGCAAGGTG TACAAATGGT CTTGATTAGA CAATTTGGAG ATATCGTTGT	3420
	ACAACATGTT GATCAACTAA CATTACAACA TGCATGTGAA CAATTAATAA CATATTTTAA	3480
	GTAGGTGAAT GAAATGGTAA ATGAACAAAT CATTGATATT TCAGGTCCGT TAAAGGGCGA	3540
25	AATAGAAGTG CCGGGCGATA AGTCAATGAC ACACCGTGCA ATCATGTTGG CGTCGCTAGC	3600
	TGAAGGTGTA TCTACTATAT ATAAGCCACT ACTTGGCGAA GATTGTCGTC GTACGATGGA	3660
	CATTTTCCGA CTGTTAGGTG TAGAAATCAA AGAAGATGAT GAAAAATTAG TTGTGACTTC	3720
30	CCCAGGATAT CAATCTTTTA ACACGCCACA TCAAGTATTG TATACAGGTA ATTCTGGTAC	3780
	GACAACACGA TTATTGGCAG GTTTGTAAAG TGGTTTAGGT ATTGAAAGTG TTTTGTCTGG	3840
35	CGATGTTTCA ATTGGTAAAA GGCCAATGGA TCGTGTCTTG AGACCATTGA AACTTATGGA	3900
	TGCGAATATT GAAGGTATTG AAGATAATTA TACACCATTA ATTATTAAGC CATCTGTCAT	3960
	AAAAGGTATA AATTATCAAA TGGAAAGTGC AAGTGCACAA GTAAAAAGTG CCATTTTATT	4020
40	TGCAAGTTTG TTTTCTAAGG AACCGACCAT CATTAAAGAA TTAGATGTAA GTCGAAATCA	4080
	TACTGAGACG ATGTTCAAAC ATTTTAATAT TCCAATTGAA GCAGAAGGGT TATCAATTAA	4140
	TACAACCCCT GAAGCAATTC GATACATTAA ACCTGCAGAT TTTCATGTTT CTGGCGATAT	4200
45	TTCATCTGCA GCGTCTTTTA TTGTTGCAGC ACTTATCACA CCAGGAAGTG ATGTAACAAT	4260
	TCATAATGTT GGAATCAATC CAACACGTTT AGGTATTATT GATATTGTTG AAAAAATGGG	4320
	CGGTAATATC CAACTTTTCA ATCAAACAAC TGGTGCTGAA CCTACTGCTT CTATTCGTAT	4380
50	TCAATACACA CCAATGCTTC AACCAATAAC AATCGAAGGA GAATTAGTTC CAAAAGCAAT	4440
	TGATGAACTG CCTGTAATAG CATTACTTTG TACACAAGCA GTTGGCACGA GTACAATTAA	4500
55		

	AAACTTGTTA GGGTTTGAAT TACAACCAAC TAATGATGGA TTGATTATTC ATCCGTCAGA	4620
	ATTTAAAAACA AATGCAACAG TTGATAGTTT AACTGATCAT CGAATAGGAA TGATGCTTGC	4680
5	AGTTGCTTCT CTACTTTCAA GCGAGCCTGT CAAAATCAAA CAATTTGATG CTGTAAATGT	4740
	ATCATTTCCTA GGATTTTTAC CAAAACATAA GCTTTTAGAA AATGAGGGAT AATATAAAAT	4800
	GGAAGATATC TATAAATTAA TAGACGATAT CAATCTACAA AACTAGAAA ATTTAGACTC	4860
10	TCGTGTTAAT GAAGCAATAA CTACTGACAA CGATGACGCA TTATTTATTC TAGGAGAGAC	4920
	ACTTTACAAT TTTGGATTAA TGCCaCAAGG TTTGGAAGTA TTCCGCGTGT TATATCACAA	4980
	ATATCCAGAC GAAAGTGAAT TGCTGATTTA TTTTATTGAA GGTTTAATGT CTGAAAATCA	5040
15	AACTGACGAA GCGTTAGAAT ATTTAAGTTA TGTGAAACCA TCACCTGAAA AGTTGATGTT	5100
	AGAAGCAGAT TTATATCAAC AAATTAATAT GATGGAAGTT GCTATTGATA AATTACAAGA	5160
20	AGCACTTGAA CTAGAGCCAA ATGATCCAAT AATCCATTTT GCATTGGCTG AAATGTTATA	5220
	TTATGATGGT CAATATTTAC GTGCTACCTC TGAATACGAA ACCGTTTTAG AACTGGTGGA	5280
	ATATCAAGTT AATGGTGTA ACTTATTCTC TCGTATGGCA GATTGTAGTT TACAAAGTGG	5340
25	KAACTATAGT GATkCcGATt CgCTTATACG ATGrAATTAA TGAAGATGAA ATGACTTCAG	5400
	AAGATTATCT CAAAAGAGCC ATTTCTnACG ATAAAAATGA CATCACTCAA GAAGCAATTA	5460
	AAATAATGAC TACATTACTT TCTAAAGATC CTGATTATAT TCAAgGCTAC TTGTATTTAC	5520
30	aATCaTTATA TG	5532

(2) INFORMATION FOR SEQ ID NO: 531:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 942 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 531:

	AATTGGGTTA TACTATAGGT AAATTTAAGG AGGTAAGAAA ATGGATAAAA AAGAATTAGC	60
45	GAAATTTATA GGCAATAAAA TCAGATACTA TAGAACCaaa TTGaACTTAA CTCAAGATCa	120
	ACTTGAGAGAA AACTCmACa CTAAAAArGC TACTATTTCa AATTATGAGA CAGGGTACAG	180
	AACTCCTAAA CAAGATGATT TGTTTGAAAT TGCTCATATT TTAAATATCA GTATCGATGA	240
50	TTTGTTCCT ACAAGAAATA ATAAAAAa CGACATCACT TCCATATACA ACAAACtCAC	300
	ACCTCCCCGC CAAGAAAACG TACTTAACTA CGCAAATGAG CAATTAGATG AACAGAATAA	360

55

AACTGGTGCT GGCATAGGAG AAGAATTATA TGATGACATA TTGCATGAAG AAGTATTTTT 480  
 TAAAGAAGAC GAAACGCCAT CAAATGCTGA TTTTGTATT TTAGTTAATG GTGATTCAAT 540  
 5 GGAACCTATG TTAAAACAAG GAACATACGC TTTTATTAAG AAAGAAGATT CTATTAAAGA 600  
 TGGTACAATT GCACTCGTTG TATTAGATGG AGTAAGTCTT ATCAAGCGTG TAGATATATG 660  
 CGAAGACTAT ATTAATTTGG TATCTCTAAA TCCGAAGTAT GATGATATCA AAGTCGCTTC 720  
 10 GTTTAGTAAT ATTAAAGTAA TGGGCAAAGT TGTATTGTGA TTAATAGCGC CTATATGGCA 780  
 CTTTAATATA AAAGACGTCT ATTTCAGCAG TGTTTAAAG GAGTTTATAA TGAAAATAAC 840  
 TAATTGCAAA ATAAAAAAG AACTATAGT ATATGAAGTT TTAAGTAGT GTAATCAACC 900  
 15 ATTCATTAT GAGTTACCTA AAGATTTATC GTCACATAAT GC 942

## (2) INFORMATION FOR SEQ ID NO: 532:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 417 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 532:

TGGnAAATGC AAACCAAAAT ATGaTCCTCG TGcAAGTTGA AGCGGGACGT TATGAAGAAT 60  
 30 GGGTAAAGAA TGGTTATTTT AAACCGTCAG AAGATAAATC AAAAGAAACA TATACAATTG 120  
 TTATCCCGCC ACCAAATGTA ACTGGTAAAT TACATTTAGG ACATGCATGG GATACGACTT 180  
 TACAAGATAT CATTACACGT ATGAAACGTA TGCAAGGATA CGATACGTTA TACTTACCAG 240  
 35 GTATGGATCA TGCTGGTATT GCGACACAGG CAAAGGTAGA AGCTAAATTA AATGAACAAG 300  
 GAATAACTAG ATATGATCTT GGTCTGAAA AGTTTTTAGA ACAGGCATGG GATTGGAAAG 360  
 40 AAGAGTATGC GTCATTTATT CGTGCGCAAT GGGCTAAATT AGGTCTAGGT TTAGATT 417

## (2) INFORMATION FOR SEQ ID NO: 533:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 733 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 533:

GATCCTGAAC CCGCATTGT TTCCACTAAA ACAGTATGCC CACTTTCTAC TAAAGCGTGC 60

55



ATTTTCATAC CATCCACCTC CATAATCATC TTAACGCGAA CATTTTGAAA GCGCAATCAA 180  
 AAATCCACAA AATTGTAAG GTTATTACAC TGACTTTTCC GAAAATTGTG GTAAAATATA 240  
 5 ATTAAGAAAG AACAAGGAGG CACTTACTAT GATTACTTAC AAAAATATTT TAATCGCAGT 300  
 TGACGGTTCA CATGAAGCGG AATGGGCATT TAACAGAGCA GTTGGTGTG CTAAACGTAA 360  
 CGATGCGAAG TTAACAATTG TGAATGTAAT TGATTCAAGA ACGTATTCTT CTTATGAAGT 420  
 10 TTATGATGCT CAATTTACTG AAAAATCTAA GCATTTTGCA GAAGAATTAT TAAATGGTTA 480  
 TAAAGAAGTA GCTACTAACG CTGGTGTTAA AGATGTAGAA ACGCGTCTAG AGTTTGGyTC 540  
 15 TCyTAAATCT ATCATTCCTA AAAAGCTTGC ACATGAAATT AATGCAGACT TGATTATGAG 600  
 TGGTACATCA GGCTTAAATG CCGTGGAAAg ATTTATTGTT GGTTCGTAT CAGAATCTAT 660  
 CGTTCGTCAT GCGCCATGTG ACGTGTTAGT TGTTCGTACT GAAGAGTTAC CAGCAGACTT 720  
 20 CCAACCACAA GTT 733

(2) INFORMATION FOR SEQ ID NO: 534:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 6060 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 534:

TGATGATCCT GAAGCGCTAT TGGATAATTA CAACACTGAA GATGTTGATG CACACAATTA 60  
 35 CAATAATATA AATCATGTTA TTTTGCCTG CGATGCGGGT ATGGGTCTA GTGCAATGGG 120  
 TGCArGCATG TTACGTAATA AATTTAAAAA GGCGGGCATT AATGATATTA CAGTTACAAA 180  
 TACTGCGATT AATCAATTGC CAAAAGATGC TCAATTAGTT ATCACTCAGA AAAAATAAC 240  
 40 TGATCGTGCT ATTAAACAAA CACCAAATGC CATCCATATT TCAGTGGATA ATTTCTTAA 300  
 TTCACCAAGA TATGAAGAAC TTTTAAATAA TCTAAAAAAA GATGATCAAG CATAATAATT 360  
 AAATAAATTA AAAAATGGAG GATACCGCCA TGTTATTGAG TACACGTGAA AAAGAAATGA 420  
 45 TAGCCCTAtT GATTAAGTAC CACGGtCAAt ATATCACTAT ACACGACATT GCTCAGCAAC 480  
 TTGCGGTGTC CTCTCGTACT ATTCACCGTG AATTAAAAGG TGTTGAAGCA TATTTAACTT 540  
 CATTTTCATT AACTTTAGAA CGCGCAAACA AAAAAGGGcT ACGCATTGCT GGCACAGATT 600  
 50 CTGATTTAAA CGATTTGAAG CAATCGATTG CACAACATCA AACCATTGAC TTATCTGTTG 660  
 AAGAGCAGAA AGTAATTATT ATATACGCTT TGATACAAGC CAAGGAGCCA GTTAAACAA 720

55

	TAGAGCTTGA	TTAAATAAG	TACCAACTAT	CTTTATCTCG	AAAGCGTGGC	GAAGGCATTT	840
	ACTTGGTAGG	TACTGAATCA	AAGAAACGTG	AATTTTTTAAG	TCAATTAAATG	GTGAATAACT	900
5	TAAATAGTAC	TAGCGTTTAT	TCAGTAATTG	AAAATCATT	TGTCTTTCAT	TCATTAAATC	960
	AAATCCACAA	AGACTTTGTT	GACTTAGAGC	GCATTTTTAA	TGTTGAAAGA	CtATTaATGG	1020
	ACTACCTAAG	TGCCTTACCC	TACCAACTTA	CCGAATCAAG	TTATTTAACT	TTAACTGTCC	1080
10	ATATCGTGCT	CTCCATTTC	CGTATAAAAA	ATGGAGAGtA	TGTCGCATTA	AACGATGATA	1140
	TTTATGATT	TGTACAAAAC	ACATTTGAAC	ACAAAGTaGc	AAGCGAACTT	GCTGATAAAC	1200
	TTGGTCAAAT	ATATGACGTC	ACGTTTAATC	AGGCAGAAAT	TGCTTTCATT	ACTATCCATT	1260
15	TACGTGGAGC	TAAACGAAAA	AATCTTAATG	ATACATCATT	AAATAATCGT	TGTGAAGAAA	1320
	ACAAAATTAA	AGCGTTTGT	AACAAAGTAG	AAATGATTTC	CGGTATGACA	TTTGCAGATT	1380
20	TGGATACTTT	AGTAGATGGA	CTGACGCTAC	ACCTTAATCC	TGCAATCAAT	CGTTTGCAAG	1440
	CTAATATCGA	GACCTATAAT	CCGTTAACAG	ACATGATTAA	GTTCAAATAT	CCAAGACTAT	1500
	TTGAAAATGT	AAGATTAGCT	TTAAATGATT	GTTGGCCTGA	TTTGATTTT	CCAGAGAGTG	1560
25	AAATTGCTTT	TATAGTtTTA	CACTTTGGTG	GCTCGATTAA	AAACCAAGGT	AATCGATTTT	1620
	TAAACATATT	AGTCGTTTGC	AGCAGTGGTA	TGGGAAC TAG	TCGTCTATTA	TCAACTCGTC	1680
	TAGAGCAAGT	TTTTAGTGAG	ATTGAGCGTA	TTACACAAGC	ATCAGTCAGC	GATTTGAAGT	1740
30	CACTAGATTT	AAGTCAATAT	GATGGCATT	TTTCTACTGT	GAATTTAGAC	ATCGACTCCC	1800
	CCTATTTAAC	GGTAAACCCA	TTATTACCAG	ATAGTGATAT	CAGTTATGTC	GCACAGTTTT	1860
35	TAAATACAAA	GTCTACGTT	CAAGAGACGC	ATGATAAATC	ATCAAACATG	ATTGATAAGG	1920
	ATGATGTTCA	TGTTGAAACG	AAAGATGTTG	ATGGCAACAC	ATCTTTTGAA	AATGAACAAA	1980
	CTTCATACTT	AAC TTCAGTT	TTCGAAAAAC	ATTTAAGTGA	CGAAAAATCA	GAACAATTAT	2040
40	TGCATCATAT	GCGTTCGGGT	TTAACTTTGC	TTGATT CAGT	GAAAATAGTT	AGTACCGAAG	2100
	TTAAACAGTG	GCAAACATAT	ATCGCAGATT	ATCTATATCA	ATGCGATGTA	ATAAACGATC	2160
	CAACGTCATT	CGCTGAAC TA	CTAGAGCAAC	GATTGATTGA	CAATCCAGGA	TGGATATTAA	2220
45	GTCCATATCC	TGTTGCAATA	CCACACCTAA	GAGACAATAT	GATTAAACAC	CCTATGATT	2280
	TAATCACAGT	TTTAGAAGAA	CCGTTAACAT	TGCCTAGTAT	TCAAAATGAC	AATCAAACAA	2340
	TTAAATATAT	GATTTCCATG	TTTATTTCTG	ACAATGATT	TATGGCATCA	CTGGTAAGTG	2400
50	ACTTGTCCGA	ATTTTAAAGT	TTGAAATTAG	AATCTATTGA	TACTTTTATG	GAAAATCCAC	2460
	AGGAACTTGA	AACATTATTA	AGAAACAAAT	TTTTAGAACG	AATTAAAAAA	CAATTTATTT	2520
55							

EP 0 786 519 A2

	TAACAGCCAA AATGAAGCAA TTGAAAAAGC AGGTAAAGCC TTAGTTGATA GTGGTGCTGT	2640
	AACAGATGCT TATATTCAAG CAATGAAAGA TCGTGAGCAA GTCGTATCAA CATTATGGG	2700
5	AAATGGCTTA GCAATTCCTC ATGGCACAGA TGAAGCTAAA ACAAATGTGA TTCACTCAGG	2760
	TTTAACATTA TTACAAATCC CTGAAGGCGT TGAAGGGAT GGCGAAGTAG TTAAAGTTGT	2820
	CGTGGGAATT GCTGGTAAAG ATGGCGAACA TTTAGACTTG TTATCTAAAA TTGCAATTAC	2880
10	ATTTAGCGAA GAAGAAAATG TGGATCGTAT CGTTCAAGCA AAATCTGCAG AAGAAATTAA	2940
	ACAAGTATTC GAGGAGGCAG ATGCATAATG AAAGCAGTTC ACTTTGGTGC TGGTAACATA	3000
	GGTCGTGGTT TCATTGGTTA TATTCTgCAG ACAACAATGT TAAAGTAACA TTTGCAGACG	3060
15	TCAATGAAGA AATCATTAAAT GCTTTAGCTC ATGATCATCA ATACGATGTT ATTTTAGCTG	3120
	ATGAGTCTAA AACAACGACG CGCGTGAATA ATGtTGATGC AATTAATTCA ATGCAACCTT	3180
20	CTGAAGCGTT GAAACAAGCA ATTCTAGAAG CTGATATTAT TACAACAGCT GTTGGTGTTA	3240
	ACATACTACC TATTATTGCT AAATCTTTTG CGCCTTTCTT AAAAGAAAAA ACAAACCATG	3300
	TTAATATTGT TGCTTGTGAG AATGCTATTA TGGCAACTGA TACATTGAAA AAAGCAGTAC	3360
25	TTGATATTAC TGGCCCTCTT GGTAACnaTA TTCATTTTGC TAACTCAGCA GTTGATAGAA	3420
	TTGTACCATT ACAAAGAAT GAAAATATAT TAGACGTTAT GGTGAGCCA TTTTACGAAT	3480
	GGGTTGTTGA AAAAGATGCA TGGTATGGTC CAGAACTAAA CCATATTAAA TATGTTGATG	3540
30	ATTTAACACC ATATATTGAG CGTAAATTAT TAACTGTGAA TACAGGACAT GCATATTTAG	3600
	CGTATgCTGG tAAATTTGCA GGTAAAGCTA CAGTTTTAGA TGCAGTTGAA GATAGTTCAA	3660
35	TTGAAGCTGG CTTACGCCGT GTTTTAGCTG AAAC TAGTCA ATATATTACT AATGAATTTG	3720
	ATTTTACTGA AGCGGAACAA GCTGGTTATG TTGAAAAAAT AATAGATCGT TTCAACAATT	3780
	CTTATTTATC TGATGAAGTA ACACGTGTCG GACGAGGTAC ATTACGTAAA ATTGGCCCTA	3840
40	AAGATAGAAT TATAAAACCA TTAACATATC TTTATAATAA AGATTTAGAA CGCACTGGTT	3900
	TATTAAATAC AGCTGCATTG TTATTGAAGT ATGATGATAC AGCAGACCAA GAAACTGTTG	3960
	AGAAAAATAA TTACATTAAA GAACACGGTT TAAAAGCGTT TTAAAGTGAA TATGCTAAAG	4020
45	TTGACGATGG CTTAGCCGAT GAAATAATTG AAGCGTACAA TTCACTTTCA TAATTTATTG	4080
	AGCTTTGTTT GAAACAAGAA GTTTCCAACG TTATTGCTTA ACAATCAGTA ATAATGTAGT	4140
	AGTTCCCTTG AATTAACAAT ATTAAATTTT TGAACATAAA AAATACTCCC TTCAACATAG	4200
50	ACACTTAACT TGTGTTATGT ATGAAAGGAG TATTTTTGCG TTAATAATTT GTTTTATTTT	4260
	CGAGCCACAG CCACCTATTC AATGGCTATT GGTCATTACT AAAACAAATT CATATTAECT	4320
55		

	TTGAATAAAT TTTATTCTTC AGTTTGTTGG TCTTTCTTAG TGAATCTTCT AATTAAGAAT	4440
	GCCATACCTG CACCTAGAGC TAATTCAGCA TATGGTAAAT CGTCATTATG TGACATACCA	4500
5	GTATCTGGTA AAGTTTTAGC TTGTTGTTTA GCTTTATTAA CTTTTCCTTG TTGAGCTGAT	4560
	TTTGTCTTAG CTGGTGGTC GTCAGTGTTA GTTACATTAA GCATATCTTG ATTAGCACTA	4620
	TTGCTTCCAT TTGAAACTGT AGCTGGAGAT GCATTGGCAC CGTCGTTTTG CGTAGyTTTA	4680
10	TTGTTTGCAG CTGAACCAAC TGATTTTTGC GTATCATTAG TATCTGCTGT TGCCGTATCA	4740
	TCTTTTTGGC TAACATTAGT TGAAGTCATT TTTTCTTTTG CTTCAGAAGA TGCAGATGTT	4800
	GATGGTTTAT TCGAAACTTC AGTATCAGCT TTGCTTGGCG ATTTATCTGC TTCGTTAGAT	4860
15	GCAACGTTAG TTTCAGACTT AAGTTGTCCT GCATCAGTTT GATTTGTCGT ACTTTCTTCT	4920
	TTATCTTTTG ATGTATTAGA AGGTACATTT GGTCTGTGTTA TGTCTGCTGA AGGCAATGTT	4980
20	TCAGTTGTtG ATTCAACCAT ACTTTGATTT GTTGAATCAC TACCATCTTT TTCTGCCTTA	5040
	GCTTTATTTT CAGATTTTGG TTGTGCAACC TTGTCATTAG TTGATTGAGA TTCAGCACTA	5100
	TTATTTACTT CAGCATTTTG TTTGAATCA TTTACAGATG CATTATCTTT GCTATCAGCA	5160
25	GATGATGCTG CTTCTGTGCT CGCAGTTGTT GGAGCCGTTG CTGTTGATCC TGTGTTGCA	5220
	TTCTCGTTTG TTGCTGTAGT TGTACTATTG TTATTTGTTG TGCTTTCTGC TGGCGTTGCA	5280
	TTATCAGTTT CTGTTACAGG TTTATCAGTT GTGCCGTTAT TAGTTGATTC TACTTCTGGT	5340
30	TTACTAGTTA CATCGTTATC CATTGTGCGA CTGTTTGTTG ATGCATCTAC ACTAGAATTG	5400
	TTATTAGCTT GCGGTTTATC ATTTGCATCA TCAGTTGCTG ATGTTGCTGT TGTTTCACCT	5460
35	GTTGCCGCAT CACTATTATT TGGTGTGTGTC GGAGAAGCGT CTGCTTTGCC ATTAGCTGTC	5520
	GTCTCAGATA CGTTAGGTTG TCCAGTATTT TCTGGTGTTG CATTAGCATT TGAATTTGCT	5580
	GTTGCATCAT TATTATCTAT ACCATTATTA GTATCATTAG CATCTGGATC ATTCTGAGGC	5640
40	ACAATCGCTT CAATTGCAGG TATCGTTACA TTTTGTAATT CAGCAACTTC TGCATTTGTT	5700
	TGTGTTTTAT CTAATTTATC AGCAAATCTG TCAAATATC TACCTAAATC CGTACGTGCA	5760
	ATTTCTTTTG CCGATGCATC TGCATCTGCA TTTTAAATTA TTTCTATTTG CTTGTAAACC	5820
45	ACTTCTCTGA TTGCTTCCAA AGCATTTTTT TTAACCTCAG GATTAATACG TTGTGCTTTA	5880
	AGTTGTTCAA GCGCACTATT TTTGACAGTA GCGATTCTG CATTGTAGT TTGATCAGAA	5940
	ATATCTTCAG TTGCTTTTGA TAAAATGTCT TCTAAAGCAT TCGTAAACGC TTCTTTTTCT	6000
50	TCAGTTGTAG CATCAGCGTT GACATTTACA CCTGCTTCAA TCTGGTCTAG TGCAGTTTCT	6060

(2) INFORMATION FOR SEQ ID NO: 535:

55

(A) LENGTH: 977 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 535:

```

10 AACAAAGCCT TCCAATTATC TCGTCGGTA GAACAAGTAT TAGCAACTTT ATCACCTACG      60
   CTAAACAGTC CTTACGATTT ATACGGCAGC ACAAAAATGC TAGATATTAC ATTTCGATTCA      120
   TTTGAACATG ATGGTACAAC GTACCCTGTC GACTATGCTA CGTTTGAAAA TGATTATGAA      180
15 GATAATAAAG ATCCTGAGTT TAGACGTAAA AGTTTCAAAT CGTTTAGCGA TGGGATTCTGA      240
   AAATATCAGC ATACTACCGC GGCTACATAT AATATGCAAG TACAACAAGA AAAAATTGAA      300
   GCTGATTTAC GTGGATTTGA ATCAGTCATC GATTATTTAT TACATAGTCA AGAAGTAACG      360
20 CGTGATATGT TTGACCGTCA AATCGATATG ATTATGCGTG ACTTGGCACC AGTTATGCAG      420
   AAATATGCTA AACTTTTACA ACGTATTCAC GGATTAGATA ACATGCGTTT TGAAGACTTG      480
   AAGATTTCTG TAGACCCTGA TTATGAACCA GAGATTTCAA TTGAAGACTC AAAAAATTAT      540
25 ATTTTCGGTG CGTTAAGTGT TTAGGTGAT GACTATACAA ACATGTTACG TGAAGCATAC      600
   GATCAGCGAT GGATTGATTT TGCACAAAAT AAAGGTAAAG ATACAGGCGC ATTTTGTGCA      660
   AGTCCATACT TTACACATTC ATATGTGTTT ATTTCTTGA CTGGTAAAAT GGCTGAAGCA      720
30 TTTGTCTTAG CACATGAATT AGGTCATGCA GGTCAATTTA CATTAGCTCA AAAACATCAA      780
   CCATATCTTG AATCAGAAGC ATCAATGTAC TTTGTTGAAG CCCCTTCTAC AATGAATGAA      840
35 ATGTTGATGG CCAATTATTT ATTTAACACA AGTGATAATC CAAGATTTAA GCGTtGGGTT      900
   ATTGGCTCAA TTTTATCTAG AACATATTAT CATAATATGG tACCCmTTTA TTAGAAGCnG      960
   CTTATCCACG GGGAGTG      977

```

40

(2) INFORMATION FOR SEQ ID NO: 536:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1440 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 536:

50

```

AGACAGTGAT yGaATTTTCAT TTACAGTACA CAAATCATCG AAAAATTGGT AACATTCTTC      60
TCTATTTTCT AACGTTAgwa TTgCATCAAA CAATTCATCT AACGCTGCAC CTCGTAATTT      120

```

55

AATCTATTAA AGTGTAGCGa TTTATATTTT ATTAAATCTG AATCGTTACT TTATTTAATT 240  
 TTATGCTAAT CCAGCGCGTT CGAAAATAGT GTCAACTTGa TTCAAATGAT GTTTAGGATC 300  
 5 GAAACATTCA TCCAATTCCT CTTTGTGTAA AACACTTGTA ATAGACTCAT CTTGTTCGAT 360  
 TAATTCACGG AACGGTGTTC TCGTTTCCCA AGATATCATC GCTTTTGGTT GTACTTTGTC 420  
 GTATGCTTCT TCACGAACCA TACCTTTATT AATTAATGCT AATAAGACAC GTTGTGAGAA 480  
 10 AATCAGACCA AATGTTTTAT CTATGTTATT ACGCATATTA TCTTCAAATA CAGTTAAACG 540  
 GTCCACAATA TTGTGAACG ATTCAATGCA TAATCTAGTG CATTGTAAAC ATCTGGTAAC 600  
 15 ATAATACGCT CAGCAGAAGA ATGAGAAATA TtCTTTTCAT GCCATAATGG CACATTCTCA 660  
 TAAGCTGTAG TAATATAACC ACGAATGACT CTTGAAATAC CTGTGATATT TTCAGAACCA 720  
 ATTGGATTTT GTTTATGAGG CATTGCAGAT GAACCTTTTT GGCTTTTGC AAATGCTTCT 780  
 20 TCAACTTCTC TCGTTTCGGT TTTTGAAGG TTACGTATTT CAACGGCAA TTTTCTAGT 840  
 GATGTCGCGA TTAATGCTAA TGTCGCAATA TAGTATGCAT GTCGATCGCG TTGCAATGTT 900  
 TGCGTTGATA CAGGCGCTGT GCCAATACCT AAATGTTTAC ACACATAACT TTCTATTTCA 960  
 25 GGAGGAATGT TAGCAAAAGT ACCTACTGCA CCACTCATTT TCCCTACTTC AATTTCTTCT 1020  
 CTTACTTGTT TGAAACGTTG TAAGTTACGT TGCATTTCCG TGTACCACAA TGCCATTTTG 1080  
 ACACCAAATG TAGTTGGTTC TGCATGCACT CCATGTGTAC GTCCCATCAT CAATGTATAT 1140  
 30 TTATAATTTT TTGCTTTTTC AGCTAAAACG TCGATAAATC TTTCTAAATC TTTTCAATA 1200  
 ATGTCAATTT CTTGTTTAAAT AACGAAACTT AAAGCTGTAT CTACAACATC AGTAGAAGTT 1260  
 35 AAACCATAAT GTACCACTTA CGTTCTTCAC CTAGCGTTTC AGAACTTGT CTAGTAAAGG 1320  
 CTACAACATC ATGGCGCGTT TCTGTTCAA TTTCTGTGTC ACGTTCGACA TTTACCTTTG 1380  
 CGTTTTGACG AATTTTTTGT ACGTCAGCTT TCGGTATATG TCCTAATTCA CTCCATGCTT 1440

40 (2) INFORMATION FOR SEQ ID NO: 537:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 784 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 537:

50 GATAAATCTA TnAGTTTCC GTCCAAATAT CtGCaCCTAA AGCTTTTAAg TGTCTACAA 60  
 TATCTgTATA ACCTCTATAA ATATGTTTAA CATTGTAAAT TGTAATTACA CCCTCAGCAA 120

55

CATGTAATGT TGATGGTTTT ATCGTTGCTG TGCCTTCGTC AACTTCAATA TTTGCACCCA 240  
 TGCGCTTTAA TTCTTCAACA TGTTTAAAC GCTCCGATA AATCGTATCA GTTACAAATG 300  
 5 AAGGACCATT TGCCATAAAT AATAATGGTG TAATAGGCTG TTGCAAATCA GTAGCAAAC 360  
 CTGGATATAC TAGTGTTTTA ATATCAACAA ATTGATATGG CGCATTATTA TTGATGCGAA 420  
 TTCTTTCGTC TCTTACATCA ACATTACAC CTAATTCACT AAATTTAGCA GTTAATGTTT 480  
 10 CTACATGTTT CGGAACAATA TTATTTAATA TAACATTTTC TCCACATGCT GCAGCGATAC 540  
 ACATATATGT GCCTGCTTCA ATTCTATCAG GTATAACTTG ATACTCAGAA CCATGTAATT 600  
 15 CTTTGACGCC ATTGATTTTA ATTGTTGATG TACCCGCTCC CTTAATATTA GCTCCCATAC 660  
 TTGTTAAGAA GTTAGCAACA TCAACTACTT CCGGTTCTTT AGCAGCATTT TCAATTACAG 720  
 TTTGTCCTGT TGCATAAACT GCAGCTAGCA TAATGtnAAT TGTTGCACCT ACGCnAACCA 780  
 20 TATC 784

## (2) INFORMATION FOR SEQ ID NO: 538:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 3733 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 538:

CAATCTCCAC CAATGACACG TCGTCATTCTG TGACCTCATA CCArACAAAA AACAGTCTCG 60  
 CAATCAAGAC TGTTTTCCAC TCAATATATT CATCCATTAG CGTAATAGAT TATTTGACTT 120  
 35 CTGTAGCTAC AAAGaTTTTA CGTTTTTCCC AAACGCCTGT cTTTTCATTG TAATCATCAC 180  
 AAGTAATTAA TGTTAATTGT TTATCTTTAC CTTTTTGTTT ATCTAGAACT CCTACATCTG 240  
 40 TAGGCTTAAC ATCTCTTATA CTGTGCATT TATACTTACG TGTTTCATTA CCAACTTTAA 300  
 AGTACACCAT ACTACCTTTT TTGGCTGCTT TAAGATTGT AAATTGATAG TTCGGACGGT 360  
 CAATGAAAGT GTGTCCTGCA ATTGAAATAT TTTGATCATC TAGTGATTCA TTTCTTCTG 420  
 45 CAAAGCTTAC ACCTCTATTT AATTGTTTCTG GTGTTGCTGG TCCTGGATAT ACTGGTTCTT 480  
 TAATATCAGC ATCTGGAATT TCAATATAGC CTGCCACTTT CGATTTATCT TTCGGAATTT 540  
 GAGGTTTAGC TTGCTGCTTT TTATCTTTAC TCGCCTGTTT TTTTACATTT TTATCATATT 600  
 50 GTTCAATCTT TTCATCTTTA TCTTTATCGT GAAGATAATT ATCGATATGT GGTTTAGCAA 660  
 ACAAATATGC TGCCACTAGG aTAAGTACCA CACCAGCGAT TGTCAATTAAT CGATTGTGCC 720

55

## EP 0 786 519 A2

	AGTATACCAT TAATTTCAAA ATGACTCATA GCAATTCATT TTATACTATA AAATTTACAT	840
	GTATACTTTT ACGTTAGATT TCATTACACA TATTTGCATT CAAATAACGA AACGCTTTTA	900
5	ATAATTACTA AGGGGGAATT GATATGATTA GATACGCTAA AAAAGAGGAT TTAAACGCTA	960
	TATTAGCGAT ATACAATGAT GCCATTATCA ATACTACAGC TGTTTATACT TATGAACCAC	1020
	AAACCATAGA CGAACGTGTC GCATGGTTTG AAACGAAACA ACGTAAGCAT GAGCCTATCT	1080
10	TTGTATTTGA GGAAATGGA AGTGTCTTAG GGTTCGCAC GTTCGGTTCA TTTAGACCTT	1140
	GGCCAGCATA CCTATATACA ATCGAACATT CTATTTATGT CGATGCTTCA GCTAGAGGAA	1200
	AAGGTATTGC TAGTCAATTA CTACACCATT TAATTGTGGA AGCAAAAGCT AAAGGTTATC	1260
15	GTGCGCTAGT TGCAGGCATT GATGCTTCCA ACAAAGCGAG TATTCAGTTG CATCAAAAGT	1320
	TTGCTTTTAA GCATGCCGGC ACACTGACCA ATGTAGGTTT TAAATTTAAT AGATGGTTAG	1380
20	ATTTAGCATT TTACGAATTA GATTTACAAG ACTAGTAATG TTTGAATCAC ATAATATAAA	1440
	CAAGACAACC ATGTTAATTC CCTTAACATA ACAAGCCAAC ATATAAAATT TTAACTTCT	1500
	CAGGGGAGTG GGACAGAAAT GATAAAGAGC CACTAATGAT TTATTATGTA GTGGTTCTTA	1560
25	CACATTAGCC ACAGCTAATG TGTACTTAAA AATAGGAATA CATGAGTAAA ACTCATGCAT	1620
	AAGAAATACT AATTTCTATA GAAAAAGTAT TTCTTTATCG TCGTCCCACC CCAACTCGCA	1680
	CATTATTGTA AGCTGACTTT TCGTCagCTT cTGTGTTGGG GCCCAAAAAG CTTGTTACAA	1740
30	GCGCATTTTC GTTCAGTCAA CTACTGCCAA TATACTTTG TAGAGCATAT TACATTGATT	1800
	TACATTGTCC CTTTTATTTA TTCTTTTCAA ATACTATCCC CATAGCTTTG ATTTAACGCT	1860
	TTTTCTCAAT AACAAAACGA ATATAGTAGA ACATGAAAC GATAATCATG CTGAGCGATA	1920
35	AAGATTTAAA TAATAGATTG ACCCACGTTT CCTCAGTCGT ATATCCATAT GTAATCGTTG	1980
	TGTTAATGAT GAATGCTATA AAGATGATTG ATAGTCTTAG CATATCATCA CTCCTTTTAA	2040
40	GTTATTTTAG ATATACGGGG GCGCTTTTGC AATCACTATT TTGATTAGTA TGCATTTTCC	2100
	ATAAATCTTT CAACTTCTTC AGAGATAATT AAGAAGCATC TATCTGGTAC TAATGATCCA	2160
	GACAGATGCT TCTTTTTTAT CAATATTTTA TTGTTATCTC ATTAATTATT TTTAACCATA	2220
45	TCTTCAGCTG TGCCAAAGAT TTTACGTTTA ATTGCTTCGC CAGTTGGTGT GCCTGCTAGT	2280
	CCACCCAATC CAGTTTCACG TAATGATGCA GGAAGGTTAC GACCAACCTT ATCCATTGCT	2340
	TCAATAACTT CATCAACAGG GATTCTACTT TCAATACCTG CTAATGCTAA ATCTGCTGAA	2400
50	ATTAAAGCGT TACCCGAACC AATTGCATTT CTCATAACAC AAGGAATTTT AACAAGTCCG	2460
	GCTACTGGAT CACAACTAA ACCTAATAAA TTACTTATCG CTAATGCCAT AGCGTGCCCCG	2520
55		



	GAACCAACTT CAGnTTGGCA GCCACCTGTT GCACCAGCTA CACTTGCATT GTTTGCTACG	2640
	ACACGCCCAA ACAATGCTGA AGTGAATAAG AAATCAATCA TTTGCTCTTC TGTAAATCA	2700
5	TGTGTTTTTT CTAATTTAAA AAGTGCACCG GGAATGGTAC CCGAGGAACC AGCTGTTGGC	2760
	GTtGCACAAA TAATACCCAT CGCAGCATTG ACTTCATTG TTGCAATGGC AcCtTTGcTG	2820
	CGTCAATCAT TTCATATCCA GACAAAGCAT GATGTGTTTC ATTATAATCA CGTAGTTTAG	2880
10	CAGCATCATG ACCAGTGTAG CCCGTTACAC TTTCAACCCC ATCACCTGTC GTCCCTTTGA	2940
	TTACTGCGTC TCGCATGACA TCTAAATTTT GTTTCATTG CGCTCGCACT TCATCACGTG	3000
	ATTTACCGCT TAATTCATT TCTTCTTTAA CCATGATATC CGCAAATGAC ATATTATTTT	3060
15	CTACGGCATA ATCTATAGTC TCTCTAATTG AATCAAACAT GTTTATTCCC CCTCTAATTT	3120
	ATATAGGAAA CGTTTACGTC ACTGTATTTT TCTTTAATTG TATTTAATGT TGATTCTGAG	3180
20	ATTGCTTTAT TTAATGGTAT TACAACCAAG CATTTATCTT CATCTATCTT AATAAATTCA	3240
	TCTTTACAGT CTAATTTTAT ATCGTTGATA TCATTGATGA AATGATTTAC TTGTGCTTTA	3300
	GTCATATTTT CGTCAACAAC TAAAATTGGT AATCCATGAT TTAAATCTAC TTCTAGTCCA	3360
25	TTTATATGAA TACCTTTAAT TTTAATTGTA CCACCACCGA TTGAAATACC GATAATTTCA	3420
	ATGTAGCGAC CATCATTACG AGATGATTTG ATATAAGCAC AGTTTGGATG TTGACCAATA	3480
	CTATCGCCTT CTTCTTCGAT GATATCTATT TTAATACCAT CATCAGCTGC AATTTCTAAT	3540
30	GAAGATTTAA TTCGGTTATC AAATGTTGAA TATCCATTG CTCCACCCAC AATAGCGACA	3600
	TCTGTACCAT GTCCTTGGTG TGTTTGAGCA AATGATTCAT AATAATGTAT TTCAATATTT	3660
	TTATATCTCC CAATATTGCG CGTGCTGAAT TCCCCTTTAC TGCACCAGCC GTATGAGAAC	3720
35	TTGAAGGGCC CAT	3733

## (2) INFORMATION FOR SEQ ID NO: 539:

40	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 525 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 539:

	TGGCTGTCTT CTCTATGAGT GTAGTAAGTA AGTTAACGGA TTAAACGCCA AGGCAAATAC	60
50	GTTACTATGA AACACATGAA CTCATCAAAC CTGAAAGAAC AGAAGGTCAA AAACGTCTGT	120
	TCTCACTCAA TGATTTGGAA AGATTACTAG AAATTAAATC ATTATTAGAA AAAGGATTTA	180

55

AAGAGATAAG AAAAAAGATG ATTGTAGATG CCACGCAAAA GCCTATTGGA GA<sub>r</sub>ACTTTGC 300  
 CAATAAATCG TGGTGATTTA TCCCGATTTA TTAAATAAAA TTTGGAGGAT TTTAAAATGC 360  
 5 CAAAACGTAC TTTCACTAAA GACGACATTC GTAAATTTGC AGA<sub>a</sub>GAGGAA AaTGTAaGaT 420  
 ATTTAAGATT ACAATTCAC<sub>T</sub> GATATTTTAG GAACAATTAA AAATGTTGAA GTGCCTGTAA 480  
 GCCAATTAGA AAAAGTACTT GATAACGAAA TGATGTTTGA CGGTA 525

(2) INFORMATION FOR SEQ ID NO: 540:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1408 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 540:

TTGATTTGCT ACAAAGTATC TnCTCATTC TGTATCCTGA AAAATCTTTA GTGTAATAAT 60  
 GTTGTTCA<sub>G</sub>T TTTAATATTT TCAGTCATAG TTGACTACCT CCGTATATTT TGATTTAATT 120  
 25 AAGTTGTATA TTTTGATGAA CACTTATTGT TACTTGTGG CGCAAGTAGC AGTTTTTTCA 180  
 TTCTTCATAA AAGTATTCCT TATAGAATAT GAATGTTGCG ATACTTGCGA ATCCTGCAAT 240  
 TGaCCATGCT GtAGTGAAGT ATAGAAACGG CATAAGTACA ATCGCTAAGA CTGTGAAGCA 300  
 30 TAGTACTGCT ACTAGGTAGC TTTTATAAAT GTTACTCATT TTCTTTTTTC AACTCCTCCA 360  
 TTATTCTCTG GTCTGATAAG TCGTGATAAG GGAATTTTTT CctAGCTAAT TGGACTGGTA 420  
 TTCTGCCTCG TATCGCAATG TATCCTTCAT CTTCAAGCTC TTTATTCAGT TCTCTTATTA 480  
 35 TTTGCCTGC TTTGGATTTA GAAACAGATA AAATTACCgC AAGTTCTTTA GCTTGCAAAC 540  
 TATTTTTCAT CATATCTTTT CCTCCTTTAA AATAACTGTT GATTCTCTGG GTTATCTGCT 600  
 TCGTAATTAT CTGCAATAAT ACTTTTAGCG AAAAAGTCCA AACTGACCTT ATATAGGTTG 660  
 TTCATAGATT TCTTTACGTT AACCCCTTCC TCAAGTACAT AAGGCACCCT AAAATCATTT 720  
 ATAAACAGTC CGTTTTCGTC TAAAGTAACG GTTGGTAATT CAGGTTTGTT CCGTCTATAA 780  
 45 ACTTCTCCTA GTGTAGGTTT TTGCTTTTCA GCTTGTTTAG TGAAGTCGGA AAATGCCTTA 840  
 AGTAGTTT<sub>T</sub>A TTCCTGAATC AGGATCACTG TGTCGCTCAA TCGTTTCTGC TG<sub>T</sub>AGACTCT 900  
 TTA<sub>C</sub>TAAAAT CATTTCTATT GATTACAGGC TTTCTCGTAT TTCGTTCAAT CTTCCAAACC 960  
 50 TTCCACGTCA CAACTGCCAT TGTGGTGAGG AGGGTTGTTT TGTATAGTGC GTTCATTTGT 1020  
 AATTCCTCCT ATTAAGTTGT TTGTTCAATT GTGTGTGTTA TTCTTCTTCG TCTAAATCAA 1080

CGACTTCGTA AGTGTGCTCA ATCTCGCCTG cATATGTCAC AGTAAGAGTA TCTTTGTGTG 1200  
 TGTATGTTTG ACTTTTGTty TCTtTAACTG CATAAAGTGT TAATACTATA TTGTTTAGCT 1260  
 5 TTyCTTTTTTG TTCTGGTGTC ATTTACGCTC CCCCTAmATT AGCyTCATAA CCGAATTCAG 1320  
 TCATGATTC ATGTATTTTC AATCTGCCTT TTTGTGTCCA TCTAGTTTGT AAAACTGTGT 1380  
 CTTCTCTGCC ATCAGAACGC ACAATTGT 1408

(2) INFORMATION FOR SEQ ID NO: 541:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 432 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 541:

GTTCGTCAAT GATTTTTATC CGATGCTATG AGCATTatCA AATACAAAAT GCTCTCTTAA 60  
 AAGCAGTTAT TGA CTGAAAA TCTACTTCTA AGAGAGCACT TTATTTAATT ACTTAAGAAA 120  
 25 TCTTGAAATT TCAATATACG ATGTTTATGA TAAGTCGCTT ATTCATCTT TAGGCTTGTT 180  
 ATTAGTAAGT AGTTTAATAC CACTGATTAA CCATAAGCA AATGTAATTA TGTTACCACT 240  
 TATTACAGCT CCAATAATCA ACAATATACC ACTCATTTTT TTGTTTTTAG ATGCTTTAAA 300  
 30 CATACCGATT GCACCTAAAA TAATTGAAAT GATTCCAAAT ATGAATAGGG ATAAGAATAA 360  
 TACAGTGAAA ATTGCTGCTG CTGTTTCTGc ATCAACTGGG nCAACCTCAC CATTAACTGT 420  
 TGTTGGACAC AT 432

(2) INFORMATION FOR SEQ ID NO: 542:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2426 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 542:

ATAATCATGA AGTnGCTAAA nCGCCAAATA ATGATGGTTC TGGACATGTT GTGTTAAATA 60  
 AATTCCTTTC AAATGAAGAG AATCAAAGCC ATAGTAATCG ACTCACTGAT AAATTACaG 120  
 50 GAAGCGATAA AATTAATCAT GCTATGATTG AAAAATTAGC TAAAAGTAAT GCCTCAACGC 180  
 AACATTACAC ATATCATAAA CTGAATACGT TACAATCTTT AGATCAACGT ATTGCAAATA 240

GTATAAAAAG TCAACGAAAT ATTATTTTGG AAGAACTTGC ACGTACTGAT GATAAAAAGT 360  
 ATGCTACACA AAGCATTTTA GAAAGTATAT TTAATAAAGA CGAGGCAGTT AAAATTCTAA 420  
 5 AAGATATACG TGTGTATGGT AAAACAGATC AACAAATTGC AGATCAAATT ACTCGTCATA 480  
 TTGATCAATT ATCTCTGACA ACGAGTGATG ATTTATTAAC GTCATTGATT GATCAATCAC 540  
 AAGATAAGTC GCTATTGATT TCTCAAATTT TACAAACGAA ATTAGGAAAA GCTGAAGCAG 600  
 10 ATAAATTGGC TAAAGATTGG ACGAATAAAG GATTATCAAA TCGCCAAATC GTTGACCAAT 660  
 TGAAGAAACA TTTTGCATCA ACTGGCGACA CGTCTTCAGA TGATATATTA AAAGCAATTT 720  
 15 TGAATAATGC CAAAGATAAA AAACAAGCAA TTGAAACGAT TTTAGCAACA CGTATAGAAA 780  
 GACAAAAGGC AAAATTACTG GCAGATTTAA TTAATAAAT AGrAACAGAT CAAAATAAAA 840  
 TTTTAAATTT AGTTAAATCG GCATTGAATG GTAAAGCGGA TGATTTATTG AATTTACAAA 900  
 20 AGAGACTCAA TCAAACGAAA AAAGATATAG ATTATATTTT ATCACCAATA GTAAATCGTC 960  
 CAAGTTTACT AGATCGATTG AATAAAAATG GGAAAACGAC AGATTTAAAT AAGTTAGCAA 1020  
 ATTTAATGAA TCAAGGATCA GATTATTAG ACAGTATTCC AGATATACCC ACACCAAAGC 1080  
 25 CAGAAAaCGt TAACACTTGG TAAAGGTAAT GGATTGTTAA GTGGATTATT AAATGCTGAT 1140  
 GGTAATGTAT CTTTGCTTAA AGCGGGGGAA ACGATAAAG AACATTGGTT GCCGATATCT 1200  
 GTAATTGTTG GTGCAATGGG TGTACTAATG ATTTGGTTAT CACGACGCAA TAAGTTGAAA 1260  
 30 AATAAGCAT AATTATATTG GGGGAAGAGC ATCTATATAT TTTTTTAAGT ATATAAGACG 1320  
 TCTTATTTC CCTTAATTTA TTGTGAAGTA TATGCAAAAT GCAATGAATA GATTGTCCAT 1380  
 CATTTTAACG TTATAATGAA TTTAACGACT TAGAACTACA CAAGTAAAGG AGAATGAAGA 1440  
 35 TGTCTCGAAA AACGGCGCTA TTAGTTTTGG ATATGCAAGA AGGTATAGCG AGTAGGTAC 1500  
 CTAGAATAAA AAATATTATT AAAGCGAATC AGAGAGCAAT TGAAGCAGCA AGACAACATC 1560  
 40 GAATACCAGT CATTTTCATA CGTTTAGTGT TAGATAAGCA TTTTAATGAT GTCTCCTCGA 1620  
 GTAATAAAGT GTTTTCAACA ATTAAAGCTC AAGGATATGC GATTACTGAA GCAGATGCAT 1680  
 CTACACGAAT ACTTGAAGAT TTAGCACCAC TAGAAGATGA GCCGATTATT TCTAAGCGAC 1740  
 45 GCTTTAGCGC ATTTACAGGT AGTTACTTGG AAGTTTATTT ACGTGCAAAT GATATTATC 1800  
 ATTTAGTATT AACGGGTGTC TCTACAAGTG GAGCTGTATT GAGCACGGCA TTAGAAAGTG 1860  
 TAGATAAAGA CTATTATATT ACTGTTTTAG AAGATGCTGT TGGTGATAGA TCAGATGATA 1920  
 50 AACATGACTT TATTATTGAA CAAATTTTAT CACGCTCATG TGACATTGAA TCCGTAGAGT 1980  
 CATGAAAAG TAGTTTATAG TTAATATAAC GTCAATTAAA GCTCGGCAGT AATGTTTGAG 2040

55

GAGGAACATT TGAACATAAA ATAATATATT TATATAAAAC GACCgAGGCG TTCGAACTGA 2160  
 ATGtCCTCGG GTTTAATTGA ATAGAAATCG GACTTATGAA CGAAATATGT TTAAGTCGAA 2220  
 5 CTCCTTGTTT ATACTTATAA ATTTTACGGG TTAAATATAA TACTTATTTA CCTGTAATAT 2280  
 ATGCATAATT nCTTCAGTCG GTCAGCCTGT CGTTGCATAG TTCCTATGCA GCAAATGCAT 2340  
 ATCCTAATCC TTTAACATTG GCATTnCTGC AAATGAACGC ATAGAATCCA TTCACTGTTA 2400  
 10 ACTTTTTnCA ACAAATGTCT nACATG 2426

## (2) INFORMATION FOR SEQ ID NO: 543:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1874 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 543:

GAGTTGGGGA ATGTGCTCAA AATATGCGGA CTTTATGCAT TyCGGAATTG sCCaATTGCA 60  
 25 GCTTTAAGCT ATGGTCAAAA AAAGAGGGTC ACTATAGCAT CTGTTTTAGT CTTAAATCCG 120  
 GAAATAATCA TATTGGATGA ACCGACTGCT GGTCAAGATT TCTATCATTA TAATGAGATA 180  
 ATGTCATTTT TAATTGAACT AAACAGACAG GGGAAAGACGA TTATTATGAT TACGCATGAT 240  
 30 ATGCATTTAT TGTCTGAGTA TAGTTCAAGA ACAGTTGTAT TATCAAAAGG TCAAGTCGTT 300  
 GCTGATACCA CGCCAgTATT GGTTTTAAAT GATAAAAAAA TCTGTGAGAT TGCATCATTG 360  
 AGACAAACAT CGCTATTTGA AATGGCCGAA TATATAGGGA TTAGCGAGCC ACAGAAATTA 420  
 35 GTACAATTAT TTATTAACCA TGATAGGAAG GTGAGACGcC AATGAATCAA TATAATACTA 480  
 TAGGTTTTCA CCCGGGAAAT AGTCGTATTC ATCAATTAAA TGCGACTGTT AAACCTTTTAT 540  
 40 TCTTATTAGT TGTTTCTATT TCTGCAATGG TGACTTATGA CACAAGATAT TTAATTTTAA 600  
 TTAGTGCTTC ATCTATTTTA TTGGTCAAAT ATGCTCATAT TGAATGGAAA CAAGTTCGCT 660  
 TTGTTGTTAA ATTCATTCTG TTTTTCACAA TAeTAAATAT TATTGCCGTG TACATATTTG 720  
 45 ACCCTGAATA TGGTGTGAAG ATTTATAATC AGCGTACAGA GTTAGTCAAT GGTATTGGTC 780  
 GATTTACGCT AACATCACAG GAATTATTCT ATCTTTTTTA TCTAATATTA AAATATATTA 840  
 GTACAGTTCC TTTAGCGTTA ATATTTTTAT TCACAACGAA TCCGAGTCAT TTTGCTGCAA 900  
 50 GTTTAAATCA GCTAGGTGTG AATTATAAAA TCAGTTACGC AGTCTCACTA GCATTAAGGT 960  
 ATATTCCAGA TATTCAAGAA ACATATTtTA ATATTTCA CA AGCGCAACAA GCAAGAGGAT 1020

55

TACCTTTAAT ATTTTCTAGT ATCGAAAGAA TTGACACTAT TAGTACTGCT ATGGAGTTAA 1140  
 GACAATTCGG ACAGTATAAA AGGAGAACCT GGTACGTCAA AAAACAATTA AAAAAAGATG 1200  
 5 ATTATGTTGT TTTGTGTTTG ACGTTAATAC TTCTGATGTT AGTAGTTACA TTATTCTTTT 1260  
 TAAATAATAG TCGATATTTT AACCCGTGGC ATTAGTATTC ATATAAATAG TCTTTAAATA 1320  
 GAAATAGGAG GGAGACATTT AATGATAAAT ACTGAAAGAT TAAATTTAAT GATTCCAAGT 1380  
 10 TCCTCGCATT TAATTGAACT TTATAATATT TGTAGTCATC CACAAGCAAA TATATACACT 1440  
 CCCAAAGGTT TACATAATTC CAAATTAGAC ACACAACGGT GGATTGAAAA ATGGCGAAAC 1500  
 CATTGGCAAC AATATCAATT TGGTTACTTT GTATTGGTAA AAAAAATAGA TTGTAGTGTT 1560  
 15 ATTGGTATTT GTGGATATGA ATATCGACAA TTAAAGCAAG AaACAGTATT AAATTTATTT 1620  
 TATAAATTAC ATCCAAGTTT TGAAGGACAA GGTACGCAT GTGAGGCTAT TACAGCAATC 1680  
 20 ACAAATTTTG TGAATTATAT CGATCAAGAA ACAGTAAAAG TTATCAGGAC AAATAAGTGT 1740  
 AACCAACGTT CAATAAATTT AGCAGAAAGG CTTAAATTCA AGCGAGACGA TACTATGGAC 1800  
 GACATTATCA ATCAAGGAGA TATTGTGTTT TAAaAATAAA ATACTATGAC ATTATCTAAA 1860  
 25 AAATAAAATT AAAA 1874

(2) INFORMATION FOR SEQ ID NO: 544:

(i) SEQUENCE CHARACTERISTICS:  
 30 (A) LENGTH: 5280 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 544:

TCAACATTTT TAACACCAAT GTGAAAATGA TCTATGTGAT TTGCAATGGC TTGATTTGTA 60  
 40 ATATGTGTGC CTAAATGACC TGAGCACCT GTTAACATAA TATTCATTCA CTTTCATCTCC 120  
 TAATCTTTAT ATACATAACA TAATACTTAT TTGATGGTTT TCAAAACATT TGATTTTATA 180  
 AAAAATTCTA ATCTGTATTT ATTGTGACG TGTATAGTAA ATACGTAAAT ATTATTAATG 240  
 45 TTGAAAATGC CGTAATGACG CGTTTTAGTT GATGTGTATC ACTAATATCA TTGAAAATTT 300  
 TAATCaGGTA CTACGACAAT ATGATGTCTG TTTTGTGTCT GAAAGTTTTA CAGTTTTTAA 360  
 50 AATAAAAATG GTATAAAGTG TGATTTGTAT AAAAAAGAGT CTCGACGGAT AAGAATTGAT 420  
 TAATAACAGT TAGCATTTTA TTAATTACCT TAACAATGAT TCAAGTTTAG TTAAATGAGG 480  
 TTTAATTTGA AAGGGGATAG CGCCTCAATA TAATGTAGGT AGATTGTTC TATTACGTAA 540

55

	AAGCAAGTTT AGCATTGGGA ATGTTAGCAA CAGGTGTAAT TACaTCGAAT GTACAATCAG	660
	TACAAGCGAA aGCAGrAGTT AAaCAACAAa GTGAaTCAGA GTTAAACAC TATTATAATA	720
5	AmCCAATTTT AGAGCGTAAA AATGTGACTG GATTTAAATA TACTGATGAG GGTAAACACT	780
	ATTTAGAAGT CACAGTAGGG CAACAGCATT CTCGAATCAC TTTACTTGA TCTGATAAAG	840
	ATAAATTTAA AGACGGAGAA AACTCAAATA TAGATGTGTT TATCCTTAGA GAAGGTGACA	900
10	GTAGACAAGC AACAAATTAC TCAATTGGTG GCGTTACAAA ATCAAATAGT GTGCAGTATA	960
	TTGATTATAT CAATACGCCA ATTTTAGAAA TCAAGAAAGA TAATGAAGAT GTACTTAAAG	1020
15	ATTTTACTA CATTTCAAAA GAAGACATCT CATTAAAAGA ACTTGATTAT AGATTAAGAG	1080
	AACGTGCGAT TAAACAACAC GGCTTGTTT CAAATGGTCT TAAACAAGGT CAAATTACAA	1140
	TTACAATGAA TGATGGCACA ACACATACAA TCGATTTAAG TCAAAAACCTT GAAAAAGAAC	1200
20	GTATGGGTGA GTCAATCGAC GGCCTAAGA TTAATAAAAT TCTAGTAGAA ATGAAATAAT	1260
	ACTTTCTAAC AACAAAGCGC TATGTTGAAT AGTGCTTGTT ATGGAAATAT ATGGAAGTTA	1320
	AGCGACGTAC TGTGCTTAG CTTCTTTTTT TGAGGGGAAA AGTTACAAA CTCACACAAA	1380
25	CAGTCGCACC ACGCATTATC TTTTGCTTAA ATAGCTTAAT CATATTTTAT GAATAGTTAA	1440
	AAACAGGTTA ATGTGAATAT CCGAATACAG CTCCTATAAT ATGGGTGTAT GATTCAAATT	1500
	ACGTAATAAA ACAATCTAAT TATAATAGAT TGGAGCATAC AACTATGAAA ATGAAAAATA	1560
30	TTGCAAAAAT AAGTTTGTTA TTAGGAATAT TAGCAACAGG TGTAACACT ACAACGGAAA	1620
	AACCAGTTCA TGCCGAAAAG AAACCTATTG TAATAAGTga AAATAGCAA AAATTAAAAG	1680
35	CTTATTATAA TCAACCTAGT ATTGAATATA AAAATGTGAC AGGTTATATC AGTTTCATTC	1740
	AACCAAGTAT TAAATTTATG AATATCATAG ATGGTAATTC TGTTAATAAT ATTGCTTTAA	1800
	TTGGCAAAGA TAAGCAACAT TATCATACGG GTGTACATCG TAATCTTAAT ATATTTTACG	1860
40	TTAATGAGGA TAAGAGATTT GAAGGTGCAA AGTACTCTAT TGGGGGTATC ACGAGTGCAA	1920
	ACGrTaAAGC TGTCGACCTA ATAGCAGAAG CAAGAGTTAT TAAAGAAGAT CATACTGGTG	1980
	AATATGATTA TGACTTTTTT CCATTTAAAA TAGATAAAGA AGCGATGTCA TTGAAAGAGA	2040
45	TTGATTTTAA ATTAAGAAAA TACCTTATTG ATAATTATGG TCTTTACGGT GAAATGAGTA	2100
	CAGGAAAAAT TACAGTCAAA AAGAAATACT ATGGAAAGTA TACATTTGAA TTGGATAAAA	2160
50	AGTTACAAGA AGACCGTATG TCCGATGTTA TCAATGTCAC AGATATTGAT AGAATTGAAA	2220
	TCAAAGTTAT AAAAGCATAA CACATATACT TGATGACGAA ATAAGTTGAA ATTGAAATAG	2280
	AGAGGTTAAG TGACGATCAA ACGTTGCTTA ACTTCTTTTT AATGCTTAAA AATTATTTCA	2340
55		

	TTAATAATAC TTCAATAATT GTTAAAAGGG GTTTAATGTG ATTATCTTAG AACGCCATCT	2460
	ATAATGATGT TGTATGATTC AAATTACGTA AAAAGACAAT CGAATATAAT ATAGATTGGA	2520
5	GCATACAATT ATGAAAATGA GAACAATTGC TAAAACCAGT TTAGCACTAG GGCTTTTAAAC	2580
	AACAGGCGCA ATTACAGTAA CGACGCAATC GGTCAAAGCA GAAAAAATAC AATCAACTAA	2640
10	AGTTGACAAA GTACCAACGC TTAAAGCAGA GCGATTAGCA ATGATAAACA TAACAGCAGG	2700
	TGCAAATTCA GCGACAACAC AAGCAGCTAA CACAAGACAA GAACGCACGC CTAAACTCGA	2760
	AAAGGCACCA AATACTAATG AGGAAAAAAC CTCAGCTTCC AAAATAGAAA AAATATCACA	2820
15	ACCTAAACAA GAAGAGCAGA AAACGCTTAA TATATCAGCA ACGCCAGCGC CTAAACAAGA	2880
	ACAATCACAA ACGACAACCG AATCCACAAC GCCGAAAACCT AAAGTGACAA CACCTCCATC	2940
	AACAAACACG CCACAACCAA TGCAATCTAC TAAATCAGAC ACACCACAAT CTCCAACCAT	3000
20	AAAACAAGCA CAAACAGATA TGA CTCCTAA ATATGAAGAT TTAAGAGCGT ATTATACAAA	3060
	ACCGAGTTTT GAATTTGAAA AGCAGTTTGG ATTTATGCTC AAACCATGGA CGACGGTTAG	3120
	GTTTATGAAT GTTATTCCAA ATAGGTTTCAT CTATAAAATA GCTTTAGTTG GAAAAGATGA	3180
25	GAAAAAATAT AAAGATGGAC CTTACGATAA TATCGATGTA TTTATCGTTT TAGAAGACAA	3240
	TAAATATCAA TTGAAAAAAT ATTCTGTCGG TGGCATCACG AAGACTAATA GTAAAAAAGT	3300
30	TAATCACAAA GTAGAATTAA GCATTACTAA AAAAGATAAT CAAGGTATGA TTTCACGCGA	3360
	TGTTTCAGAA TACATGATTA CTAAGGAAGA GATTTCTTGG AAAGAGCTTG ATTTTAAATT	3420
	GAGAAAACAA CTTATTGAAA AACATAATCT TTACGGTAAC ATGGGTTTCAG GAACAATCGT	3480
35	TATTAATG AAAAACGGTG GGAAATATAC GTTTGAATTA CACAAAAAAC TGCAAGAGCA	3540
	TCGTATGGCA GACGTCATAG ATGGCACTAA TATTGATAAC ATTGAAGTGA ATATAAAATA	3600
	ATCATGACAT TCTCTAAATA GAAGCTGTCA TCGGAAAAAC AAGAAGTTAA GTGACAACGG	3660
40	TTTACATGTT GCTTAGCTTC TTTTATTATG CGTAATGATG TAAAAAGACG AATATTTCATT	3720
	TGTTTGTAAG AGTGGCATT CTATGTCTTA AAAGTGACGA AACTTCAAAT GTGCCAAGTG	3780
45	TTGAATCACA TCAAAATCAT TTTTATTTAA CGAACATTAT GGATTTCTTA ATTTACTTAA	3840
	CGATGATTCA AATATAGTTA AACAAGGTTT AATGTGAATG GAGCAATACG CCATCTATAA	3900
	TAAAGCTGTA TGATTCAATG AATGTAATCG AACAAATCTA ATAATTACGA ATGGAGCATA	3960
50	CAACTATGAA AATAACAACG ATTGCTAAAA CAAGTTTAGC ACTAGGCCTT TTAACAACAG	4020
	GTGTAATCAC AACGACAACG CAAGCAGCAA ACGCGACAAC ACTATCTTCC ACTAAAGTGG	4080
55	AAGCACCACA ATCAACACCG CCCTCAACTA AAATAGAAGC ACCGCAATCA AAACCAAACG	4140



CGCCTTCAAC TAAAGTGACA ACACCTCCAT CAACAAACAC GCCACAACCA ATGCAATCTA 4260  
 CTAAATCAGA CACACCACAA TCGCCAACCA CAAAACAAGT ACCAACAGAA ATAAATCCTA 4320  
 5 AATTTAAAGA TTAAAGAGCG TATTATACGA AACCAAGTTT AGAATTTAAA AATGAGATTG 4380  
 GTATTATTTT AAAAAAATGG ACGACAATAA GATTATATGAA TGTTGTCCCA GATTATTTCA 4440  
 TATATAAAAT TGCTTTAGTT GGTAAAGATG ATAAAAAATA TGGTGAAGGA GTACATAGGA 4500  
 10 ATGTCGATGT ATTTGTCGTT TTAGAAGAAA ATAATTACAA TCTGGAAAAA TATTCTGTCTG 4560  
 GTGGTATCAC AAAGAGTAAT AGTAAAAAAG TTGATCACAA AGCAGGAGTA AGAATTACTA 4620  
 AGGAAGATAA TAAAGGTACA ATCTCTCATG ATGTTTCAGA ATTCAAGATT ACTAAGAAGC 4680  
 15 AGATTTCCTT GAAAGAACTT GATTTTAAAT TGAGAAAACA ACTTATTGAA AAAAATAATC 4740  
 TGTACGGTAA CGTTGGTTCA GGTAAAATTG TTATTAAAAAT GAAAAACGGT GGAAAGTACA 4800  
 20 CGTTTGAATT GCACAAAAAA TTACAAGAAA ATCGCATGGC AGATGTCATA GATGGCACTA 4860  
 ATATTGATAA CATTGAAGTG AATATAAAAT AATCATGACA TTCTCTAAAT AGAAGCTGTC 4920  
 ATCGGAAAAA CAAGAAGTTA AGTGACAACG GCCTACATGT TGCTTAGCTT CTTTTGTTAT 4980  
 25 GTTCGATGAT TTGAGAACCC GAATTTTCGA TGGGTCCAAA TATGACGTGG AAGAGACCTG 5040  
 AATTTATCTG TAAATCCCTA TCTATCGGGT GTGAAGCACA ACGGGATCAG TTTTATTTAA 5100  
 CGAACATTAT AGATTCCCTA ATTTACTTAA TAATGATTCA ATGATTATTA AACATGGTTT 5160  
 30 AATGTGAAAG GTCAAATACG CTAATATAA TAAAGCTGTA TGATTCAATA GACGTAAGCG 5220  
 AACAAATCTA ATAATTACGA ATGGAGCATA CAATATGAA AATGACAGCA ATTGCGAAAG 5280

35 (2) INFORMATION FOR SEQ ID NO: 545:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 886 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 545:

45 AGTAAATTG CCGGTATGAT GGACACAAAC GGTGACCTTG GTCAAGGTGA ATTAGCGATT 60  
 AATCCACCTA AATCAGATT gAACGAATTA CCTTGGGCTA CACGTAAAAA TAAACAGCCA 120  
 50 GCTTCATCCG AAAAAGGTTT AAGTGGTCAT CATGGTAATG CAGCAATGCC TCAAACCAmA 180  
 TTAGATTATC AAATATCTAT TGATAAGGTC GTTGAACAGG CGCAAAAAGC TGGTATTAAA 240  
 AAGCCGTTTT CAATCGTATA TCCAAGTGAT AAAAATGGTA CCTTTATTGT ATCTAATACT 300

55

GATCAATATA GCGGTAAAAA GCTAGGTACG ATTAAATATG ATGACTACCG TATTATTGCT 420  
 AAATGGTTTA CATGGGGCAT TCCGCTTCAC GAAGGTCATT TATTCGGCAT TTAAATAAA 480  
 5 ATCATTAAAT TATTTGTATG TATCGCTTTA TTAGTAGCCA TTGGCATGGG GTTTGTCTCT 540  
 TGGATAAAGC GTACAAAAAA TACTGCAGTA AAAGTACCAC ATCGCGTAAA AAAACCAGCA 600  
 TCTATATCAC TCATAATATG TTAAATTGTA TTAGGATTAT TAATGCCATT ATTTGGATTA 660  
 10 TCACTTATCC TTGTATTTAT AATTGAATTA ATATTATATA TTAAAGATCG TCGTGCTAAA 720  
 CAATAATGCA CTAAAGTTT TGAAGTACG AAATTTACAA AATGgATTCT CGTCTCTCTA 780  
 15 ATTACTTAAA ACGGGgTtCy AaTAATAAAT CgTACTGaTG GgAAAGTTTT TACTTTTTAT 840  
 CTGtCCGaT TTTTnGAAwT TGAAGATAAA AAAGCATCTA AAACGC 886

(2) INFORMATION FOR SEQ ID NO: 546:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4336 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 546:

30 GGCAATTTGTG TCCTTATATA AGGAAGTGTG tTAAATACAT TACTGTTGTT AAGTTGTTTT 60  
 TGTAATTCAA AGAGCAGAAC AGAGTAACAT CATCAGTTGT AGTAAACGAT AATCCGGTAA 120  
 AACAACTAAA TGAAATAATG AAAGTCATTT AACCTGAACA TTAAATATA TTTGTTTTTC 180  
 35 ATTAAGAATA ATTCAAGTAT ATTTAAATCG AGGTTAATTA TCGTATGAAA CGATGCACGT 240  
 TATAATAAAA ATGTATGATT CAAATTACGT AATGAAAACA ATCCAATATA TTAAGATTGG 300  
 AGCAAATAAA TATGAAATTT ACAGCATTAG CAAAAGCGAC ATTAGCTTTA GGAATTTTAA 360  
 40 CAACAGGAAC TTTAACAACA GAAGTTCATT CAGGTCATGC AAAACAAAAT CAAAAGTCAG 420  
 TAAATAAACA TGACAAGGAA GCATTATACC GATACTACAC TGGAAAGACT ATGGAAATGA 480  
 AAAATATTAG TGCTTTGAAA CATGGTAAAA ACAACTTACG TTTTAAGTTT AGAGGTATTA 540  
 45 AGATTCAAGT TTTACTGCCT GGAAATGATA AAAGTAAATT TCAACAGCGT AGTTATGAGG 600  
 GGTTAGATGT TTTCTTTGTT CAAGAAAAAA GAGATAAGCA CGATATATTT TATACTGTTG 660  
 50 GTGGTGTAAT ACAGAATAAT AAAACATCTG GAGTTGTCAG TGCACCAATA TTAAATATTT 720  
 CAAAAGAAAA GGGTGAAGAT GCTTTTGTGA AAGGTTACCC TTATTACATT AAAAAAGAAA 780  
 AAATAACACT AAAAGAACTG GATTATAAGT TGAGAAAGCA TCTAATTGAA AAATACGGAC 840

55

	ATAACCTTGA TTTAAGATCT AAATTAAAAT TTAAATATAT GGGGGAAGTC ATAGAAAGCA	960
	AACAAATTAA AGATATTGAA GTTAACTTAA AGTAAATCAT TACGAATAAT TAAAAGTAAT	1020
5	TGAAGCGGCT TAACGGTGAA ATGTAAATTG GTGCGCATAG CTTATACAAA AAGGATGCAT	1080
	CAATCGATAT CGTCGTTAAG CCGTTTTGGT TTGTGTGTCA TGAATCCTAT CCCAATCTCC	1140
10	ATAAAGGTAA AATTTCCACC ACCAACATCA AAATTCTCCA CATCGCAACA TAACCAAATG	1200
	TTATAATAAA TCTATTACAC AAAGAGATAA ATTACTTATT CAAAGGCGGA GGAATCACAT	1260
	GTCTATTACT GAAAAACAAC GTCAGCAACA AGCTGAATTA CATAAAAAAT TATGGTCGAT	1320
15	TGCGAATGAT TTAAGAGGGA ATATGGATGC GAGTGAATTC CGTAATTACA TTTTAGGCTT	1380
	GATTTTCTAT CGCTTCTTAT CTGAAAAAGC GGAACAAGAA TATGCAGATG CCTTGTCAGG	1440
	TGAAGACATC ACGTATCAAG AAGCATGGGC AGACGAAGAA TACCGTGAAG ACTTAAAAGC	1500
20	AGAATTAATT GACCAAGTCG GTTACTTCAT TGAGCCAGAA GATTTATTCA GTGCGATGAT	1560
	TCGTGAAATT GAAACGCAAG ATTTGATAT CGAACACCTG GCGACGGCAA TTCGTAAAGT	1620
	TGAAACATCA ACATTAGGTG AAGAAAGTGA AAATGACTTT ATCGGTCTGT TCAGCGATAT	1680
25	GGATTTGAGT TCAACGCGAC TAGGTAACAA TGTCAAAGAA CGTACTGCTT TAATCTCTAA	1740
	AGTCATGGTT AATCTTGACG ACTTACCATT CGTTCACAGT GACATGGAAA TTGATATGTT	1800
30	AGGTGATGCA TATGAATTCC TAATTGGGCG CTTTGCGGCG ACAGCGGGTA AAAAAGCAGG	1860
	CGAGTTCTAT ACACCACAAC AAGTATCTAA GATACTGGCG AAGATTGTCA CAGACGGTAA	1920
	AGATAAATTA CGTCACGTGT ATGACCCAAC ATGTGGTTCA GGTTCACGTG TGTTACGTGT	1980
35	TGGTAAAGAA ACACAAGTGT ATCGTTATTT CGGTCAAGAA CGTAACAATA CTACATACAA	2040
	CTTAGCACGC ATGAATATGT TATTACATGA TGTGCGTTAT GAGAACTTCG ATATCCGTAA	2100
	TGATGACACA TTGGAAAACC CAGCCTTTTT AGGCAATACA TTTGATGCGG TTATTGCGAA	2160
40	CCCACCGTAT AGTGCGAAAT GGAATGCAGA TTCAAAGTTT GAAAATGACG AACGATTGAG	2220
	TGGTTACGGC AAACCTTGCGC CTAAGTCTAA AGCAGACTTT GCCTTTATTC AACACATGGT	2280
45	ACATTACCTA GACGATGAAG GTACCATGGC CGTTGTACTC CCACATGGTG TATTATCCG	2340
	AGGTGCTGCA GAAGGTGTCA TTCGTCGTTA TTTAATTGAA GAAAAGAACT ACTTAGAAGC	2400
	TGTGATTGGT TTGCCAGCGA ATATTTTCTA TGGGACAAGT ATTCCAACAT GTATTTTAGT	2460
50	ATTTAAAAAA TGTCGCCAAC AAGACGACAA CGTACTATTT ATCGATGCAT CCAATGATTT	2520
	TGAAAAAGGA AAAAATCAAA ATCATTTAAG CGATGCCCAA GTCGAACGTA TTATAGACAC	2580
55	ATATAAGCGT AAGGAAACAA TTGATAAATA TAGCTACAGC GCGACACTAC AAGAGATTGC	2640

EP 0 786 519 A2

	GATTGATTTA GATCAAGTCC AACAAAGATTT GAAAAATATC GATAAAGAAA TCGCAGAAAT	2760
	TGAGCAAGAA ATCAATGCAT ACCTGAAAGA ACTTGGGGTG TTGAAAGATG AGTAATACAC	2820
5	AAAAGAAAAA TGTGCCAGAA TTGAGGTTCC CAGGGTTTGA AGGCGAATGG GAAGAGAAGC	2880
	AGTTAGGGGA TCTTACAGAT AGAGTAATTA GGAAAAATAA AAACCTAGAA TCGAAAAAGC	2940
	CTTTAACAAT ATCCGGACAG TTAGGTTTAA TTGATCAAAC AGAATATTTT AGTAAATCAG	3000
10	TTTCGTCGAA AAATCTAGAA AATTATACAC TAATAAGAA TGGAGAATTC GCGTATAACA	3060
	AAAGTTATTC TAATGGATAC CCATTAGGGG CTATTAAAAG ATTAACTAGA TATGATAGTG	3120
15	GTGTATTGTC CTCTTTGTAT ATTTGTTTTT CTATTAAAAG TGAAATGTCT AAAGACTTCA	3180
	TGGAAGCATA TTTTGATTCTG ACACACTGGT ATAGAGAAGT TTCTGGAATT GCAGTTGAGG	3240
	GTGCAAGAAA TCACGGATTA TTAAATGTTT CTGTGAATGA TTTTTTTACT ATTCTAATTA	3300
20	AATATCCAAG TTTAGAAGAA CAGCAAAAAA TAGGCAAGTT CTTCAGCAA CTCGACCGAC	3360
	AAATTGAATT AGAAGAACAA AAGCTTGAAT TACTTCAACA ACAGAAAAAA GGCTATATGC	3420
	AGAAAATTTT CTCACAGGAA CTGCGATTCA AAGATGAGAA TGGTGAAGAT TATCCAGATT	3480
25	GGGAAAATAG CAAAATAGAA AAATATTTAA AAGAGAGAAA CGAACGTTCT GACAAAGGGC	3540
	AAATGCTTTC AGTAACTATA AATAGTGGCA TTATAAAATT TAGTGAATTG GATAGAAAAG	3600
30	ATAATTCAAG TAAAGATAAA AGTAATTATA AAGTAGTTAG GAAAAATGAT ATTGCATATA	3660
	ATTCTATGAG AATGTGGCAA GGGGCTAGTG GTAAATCAAA TTATAATGGG ATTGTTAGCC	3720
	CTGCATATAC TGTGCTTTAT CCAACACAAA ATACTAGCTC ATTATTTATT GGATATAAGT	3780
35	TTAAACACA TAGAATGATT CATAAATTTA AAATTAATTC ACAAGGATTA ACATCAGATA	3840
	CATGGAACCT AAAATATAAA CAATTAmAAA ATATAAATAT AGATATACCT GTATTGGAGG	3900
	AACAAGAAAA GATAGGTGAT TTCTTTAAAA AAATGGATAT ATTGATAAGT AAACAGAAAA	3960
40	TGAAAATTGA AATATTAGAA AAAGAGAAAC AATCCTTTTT ACAAAAAATG TTCTTATAAC	4020
	TTTGATAAAT ACATAGATTG CATAAGAATA AAATTTGTAT AATTTAACAT AAAAGTTGTA	4080
45	AAAGTAAAGT GAATTAAAAA CGAACATTAA ATTTAGGCAC TGTGAAAGCG CAGTGTCTTT	4140
	TTTGTGTCGA AATTGTGTAC AGAATAAGTA GTTAAATAAA GATTAAGTTG AGATAAAGTG	4200
	TTATTTCGTAA ATAAAAGAGA GTAGATCGAT AGGAATTGAA TGATATTAGT TAACTATTTA	4260
50	TTAAATTACT TAATAATGAT TAATTTTTAG TTAAAGTAAG TTTAATGTGA AGCACGACCA	4320
	TTGCTCATT TAATGA	4336

(2) INFORMATION FOR SEQ ID NO: 547:

55

(A) LENGTH: 487 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 547:

10 TAAGCTATCT GAGATAATTG CTGATAACAT TAAACCGGCA ATTTTCAGGTT TAATTTCAAA 60  
 GCCACGTTCT CTAAACATTT TGTATAAAAT TGTAGCTGTA CAACCAACTG GTTCAGCACG 120  
 ATAACATAAA GGACCAGCAG TTTCGAAATT TGCAATTCTG TGATGATCAA TTACATGCTT 180  
 15 AATTGTAGCA GAGGCAATCG TATCAGAACT TTGTTGGAAT TCGTTATGAT CAACTAAGAT 240  
 AACATCTTGA CCATCTAAAT CATCTGTTAA TAATTCCGGA GCAGGTACAT TAAATGTATC 300  
 TAACGCGAAT TGAGTTTCTG CACTCACATC ACCTAAACGG TATGCTTTGG CTCCTGAATT 360  
 20 ACCTCGAAGT TGTTCaAATT CTGCCaTAAT AATCGCAGAT GAAATTGCAT CAGTGkCTGG 420  
 aTTCTTATGT CCGAAAATAT ATGTTTTAGC CAaTGTCAAA TATCTCCCTT GTAAATTGTA 480  
 25 TTCTTTA 487

(2) INFORMATION FOR SEQ ID NO: 548:

(i) SEQUENCE CHARACTERISTICS:

30 (A) LENGTH: 871 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 548:

TTGGTGGTGC AGCAGTTATA GCAATTGTTT TAGCATTTCAT TGGTAAGTTC ACTGCATTAA 60  
 TTTCTTCTAT ACCTACmCCA GTTATGGGAG GAGTATCTAT ATTACTTTTC GGTATTATTG 120  
 40 CAGCAAGTGG CTTAAGAATG TTAGTTGAAA GCAAAGTAGA TTTGCGAAC AATCGAAATT 180  
 TAGTTATAGC TTCTGTAATT TTAGTTGTAG GTATCGGTAA TTTAGTATTT AACTTAAAAG 240  
 45 AAATTGGTAT CAACCTTCAA ATTGAGGGGA TGGCATTAGC TGCACTTTCA GGAATTATTT 300  
 TGAACCTAAT CTTACCTAAA GAGAAAAAAC AAAACAATTA AGATTTACAA ATTAAGGAGG 360  
 GCGCTTTTAT GAATCATTTA TTATCAATGG AACATTTATC TACAGATCAA ATATACAAAC 420  
 50 TTATCCAAAA GGCAAGTCAA TTTAAATCTG GTGAACGTCA ACTACCAAAC TTTGAAGGGA 480  
 AaTATGTGCGC AAATTTATTC TTTGAAaATT CTAaTCGrAC AAAATGTAGT TTTGAAATGG 540  
 CAGAACTTAA GCTAGGGTTA AAAACGATTA GCTTTGAAAC ATCAACATCA TCTGTTTCAA 600  
 55

TCATTAGACA TCCGTTTAAAT AACTACTATG AAAAATTAGC GAATATTAAC ATCCCAATTG 720  
 CGAATGCTGG TGATGGTAGT GGACAACATC CAACACAAAG TTTACTTGAT TTAATGACGA 780  
 5 TATATGAAGA ATATGGATAT TTTGAAGGCT TGAATGTATT GATTTGTGGA GAcATTaAAA 840  
 ATTCACGTGT CGCACTAGTA ATTACCAaAG T 871

## (2) INFORMATION FOR SEQ ID NO: 549:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 549:

20 TCAATCTAAT ATATACTTCA TGACTTCCCG CCAAAGGCGC AATCGTAGGA TTAGTCTGTT 60  
 TATCGATAAG ATCTATTAAT ATTGTTTCTA CTTTAGATTC ACCTATTCCC GCAAATCTTA 120  
 ATAGTTCAGA ATGTATAATT CGATTATGGT TTATAAAATG TGACAACAAT TCATTTTTCA 180  
 25 CCATTGGTTG CATTTCTTTC GGTGGACCTG GTAATAAAAT AATTGTTTG TTTTCAAAT 240  
 TCACCATCAT TCCTGGAGCC ATGCCATGAT GATTTGTAA TACAGTTGAA CCTTCAATTA 300  
 CTAAAGCCTG TTGTcTATTA TTAGGTGTCA TTTCTtGTCC TTGTTCTCA aAAwAGCTTT 360  
 30 CAATATATTG AAAGAAGGCT CATCAATAAC TAAATCTnTA 400

## (2) INFORMATION FOR SEQ ID NO: 550:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1523 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 550:

ATACCTCCAG CTAGAATACC AGCGTATTTT ATAAAATACT TCCTCCATTC AACTATATCT 60  
 45 ATATTTAATT ATTTAAATTT CGTTGCATTT TCCAATTGAA AACTCATTTT AAAATCAAAA 120  
 CTCTAAATGT CTGTGTATTA CTTAAAATTA TACATATTTT GCTTATATTT TAGCATATTT 180  
 TGTTTAAACC TATATTACAT TATATCAGAC GTTTTCATAC ACAAATAATA ACATACAAGC 240  
 50 AAACATTTTCG TTTATTATTT ATATCACTTA ACTAATTAAT TTATAATTTT TTATTGTTTT 300  
 TAAGTTATCA CTTAAAAATC GTTTGGCaAA TTCGTTGTGA CGCTTGTCaA TCTTCTAATG 360

TTTGTTTTAA TGCATCAATG AGTGCTGTTT GATTTTCAAC AATTGGACCT GGCAACTCTT 480  
 TTTTATAATC CATGTAAAAA CCTCTAAGCT CATCGCCATA TTTATCTAAG TCATATGCAT 540  
 5 AGAAAATTTG CGGACGCTTT AATACACCGA AGTCGAACAT GACAGATGAG TAGTCGGTAA 600  
 CTAACGCATC GCTGATTAAG TATAAATCCG AAATGTCTTC ATAAtCTGAA ACGTCTTTCA 660  
 10 CAAAATCATC ATGTTTCATCA ATACGTGTCA CAACTAAATA ATGCATGCGT AAKAAAATAA 720  
 CATAATCATC ATCCAGCGCT TGACGCAAAG CTTCTATATC AAAGTTAACA TTAAATTGAT 780  
 ATGAACCTTC TCGAATCGCT TCATCGTCAC GCCAAGTTGG CGCGTACATA ATCACTTTTT 840  
 15 TATCTAATGG AATATTTAAT CTTGTCTTAA TACCATTAAT ATATTCAGTA TCATTGCGTT 900  
 TATGTGATAA TTTATCATTT CTTGGATAAC CTGTTTCCAA AATCTTATCT CGACTAACAT 960  
 GAAATGCATT TTGAAATATC GATGTCGAAT ATGGATTAGG TGACACTAGA TAATCCCACC 1020  
 20 GTTGGCTTTC TTTTAAAG CCATCTTGGT AATTTTGAGT ATTTGTTCCCT AGCATTTTAA 1080  
 CGTTACTAAT ATCCAAACCA ATCTTTTTTA ATGGCGTGCC ATGCCATGTT TGTAAGTACG 1140  
 TCGTTTCGGG TGATTTATAT AACCAATCTG GTGTACGTGT GTTAATCATC CwCGCTTTTCG 1200  
 25 CTCTTGGCAT CGCTAAAAAC CATTCATTG AAACTTTGT AACATATGGT ACATTGTGCT 1260  
 GTTGAATAT GTGTTTCATAT CCTTTTTC CAACCCATAT TAATTGGGCA TCGCTATGTT 1320  
 CAGTTAAGTA TTCATATAAT GCTTTGGGGT TGTCGCTGTA TTGTTTACCA TGAAAGCTTT 1380  
 30 CAAAATAAAT TAGATTCTTG TTTGGCAATT TTGATAGTAA TTTAAAAGTC GTATATATAC 1440  
 TATGTTCTAT CAATTTTTTA ATTGTATTTT TAATCATGTC GTACCTCCGA CGTGTTTTTG 1500  
 35 TAATTATATT AATATGTATG AGC 1523

(2) INFORMATION FOR SEQ ID NO: 551:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4923 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 551:

CAGTAAGAGA TTTTCTTAAT TGAAATAAT CTTACTGCTT TTTTAAATT AATTTCGAGA 60  
 50 TTCAATATTA GTTTATCTCA TTGTGGCATT AATTGATAAA ATTGTTTTAA TGTTATAAAT 120  
 CAAAGTCTTC TTCAACAATT TCAATGTCTG CATCAGATCT ATGATATGTG AAAAAGCTAA 180  
 TTCTTATGCG GTCTAAATGC TCTAAATGGT GTCGATACTC TTCGATGGCT GCAACAATTT 240

	ATGTGGACAA TAAATCTTTT TTCTGCGGAT TATCTAACTC ATAATCAACA TGTGTCACAT	360
	TATAACGTGC TTTTITAGAA AGGCTAGCTA AAATTTGCTC GTGAAAAGCA GTTAATGAAT	420
5	CTAAATCTAA TTTGATTTGT AATAGGAAAT TGTTATTAAG TAAATATAAG TCGTTTTGAT	480
	AACGCGACAA TTTGTTTAAT ACTTCATAAG CTTGTCTAGT CGTCTGAACT ACTTCTCTGA	540
	AAAGTATTTT CTTTCTATTC TGTGGTGAA TATGTTTTTT TGTAATAGGA CGTTCTTCGC	600
10	TATAGTAATC ATAAATTTTC TCTAACTTTT CGACACGTTG TTTTAAATTA TGAATATCTT	660
	GTTTAATATT ATTAAACTCC GTCGTATCAT TTAATACTAA TTTAAACCAC ATAAAAATAT	720
15	CTGAGGATAT ATTTAATGAA TTATAGTAAA TTTTGTTC AAATTTAGGT GGTAGAAACA	780
	CAAAGTTAAC TAGAGATGAA CTTATGACAC CAATCATTAC AAGTACAAAC CTGTAAAAGG	840
	CGGTAATATA GAAAGAACCG GTATGTTGTC CCATAATGAT TAATGCTGTT ACACTCGCCA	900
20	AAGTAGCAAC ATGTGCTAAA TTAAATTTAA ATAAAATAGC AATAAGTACT ATGACGGTAA	960
	CACCCATAAT GATAAAATTA TCACTAAAAA TTGTTACCAT TGTAACAGAT AGTATGGCAC	1020
	CTATAATGTT ACCCAATGCT TGATCAGAAA CTGTTTTAAA TGAACGATAA ATACTAGGTT	1080
25	GCATTGCACA ACAGCACTGA CACCAGCTAA GGCTTTCAGA CCAACATCAT CCGGTAGTAA	1140
	AGAAGCGATA GACATAGCTA AAATAATGGC TATACCAGTT TTAATAATCC GAGCTCCTAG	1200
	TCTCAAAAAT AATGACGCCC CTTTTAAGTT TATTGAATAT CTAATATTCG TATTCATTAC	1260
30	TGTTATACAC TTACTAGTTA CAAAATTCAA GCTTATTTAT AGTTGTTAAA ATAAATCATA	1320
	CATAATACTG ATAGCGATGT AAACTTTAG TCAGAGATTA AAATAGTATA AATTTGTAAA	1380
35	ATAAAAATC ACATAGTGAC ATATCAAGTT AAACGTTAAT AGTTAACGAT ATAAAATGAA	1440
	TCTACTATGT GAGCATTTC TTTATTTTAA TTCAATTAAA AATATACTTC CTTAAAAGTT	1500
	ATTTCAATTG ACTAAAAGCA TAGTCTGCAG CTTTTAAAGT TTGTTTAATA TCTTCTTCTG	1560
40	TATGTTCACT TGTTAAGAAC CAAGCTTCAA ACTTAGAAGG TGCTAAATTG ATACCTTGAT	1620
	TTAACATTAA TTTGAAAAAT TTACCGAACG CTTGCGCGTC AGAATGTTCA ACTTGATCAT	1680
	AATGTGTGAC TTTTTCATCT GTAAAGTACA ATGTTAAAGA TCCATAAATA CGATTAATTG	1740
45	TAGCTGTGAT ATTATGTTTT TCGATTAAAT TAAGTAAACC TTCTTCTAGT TGTGGCCTA	1800
	AGCTGTCTAA TTTTTCATAA ACACCGTCTT GTTCTAGTAC TTCGAGTAAT GCAATACCTG	1860
	CTTTCATAGA TAACGGGTTA CCAGCCATTG TACCAGCTTG ATATGCAGGT CCTAGAGGTG	1920
50	CTACTTGTTT CATAATATCT TGACGTCCAC CATAGCCTCC AATTGGTAAA CCACCGCCAA	1980
	CAATTTTACC AAATGCAGTT AAATCAGGGA TAACACCTAA TAAATCTTGA GCGGCACCGT	2040
55		



	AAATTTTCATT AACCTCTTCT AAAAATCCAG GTTGAGGCAT TACCATTCCA AAGTTACCAA	2160
	CAATTGGTTC TACTAATACT GCGGCAATTT CATCACCCCA AAATTCAATT GCTTCTTTAT	2220
5	AGGCGTTAAT ATCATTGAAA GGTACAGTAA TGACTTCACG TGCACGCTT TCTGGAACAC	2280
	CAGCTGAGTC TGGAGAACCG AGCTGAGATG GGCCGCTACC TGCTGCAACC AATACTAAAT	2340
10	CAGAATGGCC ATGATAAGAT CCAGCAAATT TTATAATTTT ATTTCTTTTA GTATATGCAC	2400
	GTGCAACACG AATTGTTGTC ATGACTGCTT CTGTTCCAGA ATTTACAAAG CGAATTTTCT	2460
	CAAGAGATGG AATTGCATCA CGTAATTTTT TGCTGAATTC AATTCTAAT TCAGTCGGTG	2520
15	TACCAAATAA AACACCTTTA GCAGCTTGTT CTTGAATTGC TTTAGTAATA TGAGGATGTG	2580
	CATGCCCCGT AATAATTGGA CCGTATGCTT GAAGGTAATC AATAAATTTA TTGCCATCGA	2640
	CATCATATAA ATATGCACCG TGTCTTCTT TCATAACAAC AGGTGCACCG CCTCCTACAG	2700
20	CTTTATAAGA ACGAGAAGGG GAATTGACAC CGCCTAGAAT ATATTCGTTT GAAAGTTGTT	2760
	GTAAACGTTT ACTTTCCTA AAATTCATTT ATATCAACCT CTTTTAATTT AATATTTTCA	2820
	TCTAATATCG TATCATAAAA TTATTATAAT GAAGAAAAAG GTGATTATAT GTTGCAAAAA	2880
25	GGAGAACAAAT TTCCAATATT TAAATTAGAA AATCAAGACG GAACTGTCAT TACAAATGAT	2940
	ACATTAAAAG GTAAAAAGGC GATTATATAT TTTTATCCTA GAGATAATAC ACCTACTTGT	3000
30	ACCACAGAAG CTGTGACTT TAGAGACAAT TTAGAAATGT TCAATGATTT AGATGTTGCA	3060
	GTATATGGTA TAAGCGGTGA TTCAAAGAAA AAACACCAAA ATTTTATTGA GAAACACGGA	3120
	TTGAATTTTG ATTTATTAGT AGATGAAGAT TTAAATTAG CTAAAGAAAT GCGGTATATC	3180
35	AGTTAAAAAA ATCATTGCGC AAAGAAAGTA TGGGCATTGT AAGAACGACT TTTATAATAG	3240
	ATGAACAAGG TAAAGTATTA GATGTTATCG AGAAGGTTAA GGTAAAAACA CAAATAGAAG	3300
	AACTTAAAAA CATTTTGGGG TGACATATAT GAAAGTTGTT GGGTTAAATC GTATGCGTGA	3360
40	AGTTGAAACT GAATTACAAC AACGCTTTTC AGATTTAGAT TTAAATTTT ATAAAAAGC	3420
	ATCAGAAATA CCTGAGAGCG ACTTGGCTGA TTTAGATATA TTAGTTGGTT ATGATGGCGG	3480
45	TATCAATGAG GCATTTTAC GACGTTGCCC GAATTTAAAA TGGATTGCAT GGTTTGCAAC	3540
	GGGTGTAAAT ACATTGCCGT TAGATTATAT TGCAGATCAC GGCATACTTT TAACTAATGG	3600
	AAAAGGTGTT CAAGCTAAAC AATTATCTGA ATACATTTTA GCTTTCATTT TAGATGATTA	3660
50	TAAAAAGATG AACTATCAT ATGATAACCA ACGACAACAT ATATATGATT CGAAAAAATC	3720
	TGGTAAACGC CTATCAGGAC AAACAGTTTT ATTTTAGGT ACAGGTGCAA TTGCTACTAG	3780
55	AACTGCGAAG TTAGCAAAGG CTTTAAATAT GAATTTAATT GGTCTGAGCA AGTCAGGTCA	3840

TGCTGACATT ATTATAAATG CTTTACCAGA AACGCAAGAA ACGATTCaTT TaCTAAAGAA 3960  
 AAAACATTTT GAATTAATGA AAGATGAAGC ACTTTTATA AATATAGGAC GAGGTAGCAT 4020  
 5 AGTTAAAGAA GCGCTCTTAA TAGAAGTATT AAAAAGTAA GTTATTCGAC ATGCATATTT 4080  
 AGATGTGTTT GAAAATGAAC CTTTGAAACC TAATCATGAA TTATATGAAT TGGATAATGT 4140  
 AACTATAACA GCGCATATAA CTGGTAATGA TTATGAAGCA AAGTATGACT TATTAGATAT 4200  
 10 TTTTAAAAAC AATCTAGTTA ATTTTCTCAA TAAGAATGGT CTAATTGAGA ATGAAGTTGa 4260  
 TGCTAAAAAA GGCTATTAAA TGArATCATC ATGTAAATAT TGACACGCGC GCAATACTAC 4320  
 15 AGTTATATTT aTAGTAAGTt AATaATgATT ATATAAGAA GATGGTgATA TAGATGAGTG 4380  
 TTGAATAGA ATCAATTGAA CATGAACTAG AAGAATCAAT TGCATCATTG CGACAAGcAG 4440  
 GCGTAAGAAT TACACCTCAA AGACAAGCAA TATTACGTTa TTTaATTTCT TCACATACTs 4500  
 20 ATCCAACAsC TGaTGaAATT TATCAAGCAC TTTCACCTGa TTTTCCAAAT ATAAGTGTTG 4560  
 CGACAATATA TAATAACTTA AGAGTGTTTA AAGATATTGG AATTGTAAAA GAATTAACAT 4620  
 ATGGAGACTC ATCAAGTCGA TTCGACTTTa ATACACATAA TCATTATCAT ATTATATGTG 4680  
 25 AACAAATGTGG TAAGATTGTT GATTTTCAAT ATCCACAGTT AAATGAAATT GAAAGATTAG 4740  
 CTCAGCATAT GACTGACTTT GACGTAACAC ATCATCGAAT GGAAATTTAT GGAGTTTGTA 4800  
 AAGAATGCCA AGATAAATAA TTAACTTTG GTAGTATGAC AAATTAAAA AGCGTTACTw 4860  
 30 ACTTCATATA AGTAAGCGTA ATATTAAAGA nGTTAAACGA CATGaAAGTt GTTTAACTTT 4920  
 TTT 4923

35 (2) INFORMATION FOR SEQ ID NO: 552:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 917 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 552:

45 TCCGGCTTTA AAAACTTTTC CCAATTCAG CTTGGGCCTT TGGCATTAA ATTAACCTCC 60  
 TGGTTCGGAT TAATTGGGAA CCTAACnTT TAGGCAATAA TTGGTTTAGG CAACTTCCAA 120  
 ATTGGTGGTT CAACCAACGT CTTTGATAC CyTGcTCATT TAATTCTAAA ATGgTyrGAA 180  
 50 CGCATTTTGG TACCCAAAAt GgTGACGTTT GTTGCACGG TCTAATAAAT TGTCTAAGTT 240  
 GTCGATTGGT TTCATTAAAT GAaCACCATT TTGCTCTTCA GCAAGACCTT TGTCTACTTT 300

55

GCGATCCATT GTTTGTTCTGA AAAGAATAGC ACCCAAAATT TTATCTGGTG AGAATGAAGG 420  
 TGAAGTTACC ACACGTGTAC GCATATCGTG AACAAATTGG AACATTTCTG CTTCATTGCT 480  
 5 ATATTGATCT TCGTTTACAC CATATTCTTT AAGTGCTTTT GGTGTACTAC CACCACTTTG 540  
 GTCTAATGCG GCAATAAAGC CTTTTCCATT TTTCATTTTT TCTAATTGCT CTTTATTCAT 600  
 ACTTTCCACT CCTTAACTTT TCAATACACC TCCAGTATGA TAAAAATGAG AACATTTCTC 660  
 10 AAGTCATAAA CCTTGAAAAG TGTATAAAAT GTGAAAAATA ATTGTCAGTT TAATTAAAAA 720  
 TATTATTTTA TTCTAGGTAT GACTAACGCC ATTAATGACA TAAAGAAAAT ATGTGTAATA 780  
 15 ATCCAACCGA TTAATTCTGT CACACTAAAT TGAAAAATTG GACGTTGCGC AATAAATACT 840  
 AAAAAGGGAT ACAATGCTAT AAATAAGAAA AATAAAGGGA TATAACATAG ATAGTAAAGC 900  
 CTTTTAGAAG TATGAAA 917

20 (2) INFORMATION FOR SEQ ID NO: 553:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 432 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 553:

ACTCCCATAT CCACTGGCAA GATTTATAAT TTGTAATTCT AAAGTATAAT AGCCTATATT 60  
 TTTCAAAATC TCTTTTTTTA TGATTAACAT CCCTCTAAGT GCACCTAAAG GTTTTTATGT 120  
 35 AATGTATTAA ATATTTCCAT TTATACATAT ACATCTTGTA TAAAAAGAAA GAACTCCATA 180  
 TATACTCAAA AGGTATACTG AAGCTCTATA TATTATATAA AATTATTATA CTATTTTGAA 240  
 AAACATAAGT AAATCACTAA ACATGATTTT TTTCACTTTT AGAAAACCTT TAATACTATA 300  
 40 AAAGCACCCA CTCAGTCACT AGTTTGGGCA GTTATTGTAT GCCTATTGaa CTCAATGCgT 360  
 ATATTACAAT ACCTTTTtCG CATATTCATA TAAGacTTTG CATCTTTAAG CTTAATTGCT 420  
 ATCTCTTTCT CT 432

45 (2) INFORMATION FOR SEQ ID NO: 554:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1374 base pairs  
 50 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

TTTTGCTTTA TGTAAATGG ATTATTaTTA GAATAAAAAA TCGGTGATGA GCTAAAAAAG 60  
 TGTGTAGGAT GTTTTCCKAA CCCAATTTTT ACATCCGAAG ATATCGAACA ATATCTTCCT 120  
 5 ACTTCTACAT TATTAAAATC ACTACCAAAT CCAATATAAC TGTATTCACC AATGTGAGAA 180  
 TTCTGTATT TACACCATCT ATCTATATAG TTATTGCCAT CAAATTTTGA GTTTGTAATA 240  
 TACGCCAAGC GATGAATCTT AACATTTCGAT TCTTTAGAGG ACTGGTTTTT CAGCAAACCA 300  
 10 ATTATCTTTT CAATCGCTAT CCTCATCGTC ATTTCTCCA AGTATTTTGA TTGATAATAT 360  
 CTTTATAGCT TTGAATAATT TTAACCTACCT TTGTCGAAAC GTTAGTGTCT TTATAATCAA 420  
 TAGCATCAAT CATCGGTTTCG TTATTGTTTT GCATCTCTCT TGCTAGTTCA ACGGATTGGA 480  
 TTAGATTGTT ATAGGTAATA CCACCTACAA TAACCGTACC TTTATCTAGT ACTTCCGGTC 540  
 TTTCTGTGGA AGTTTGAATA AGGACACCAG GGAACCTCAA AATAGACGAC TCTTCTGACA 600  
 20 ATGTTCCACT ATCTGATAGC ACAACAAATG CATCTTTTTG CAATGCATTA TAATCAAAGA 660  
 AACCAAATGG CTTTAACTGT TTAACCTAATG GATCAAATTC AAATTTACTT TCTTCAATTT 720  
 TCTTCCAAC TCTTGATGC GTTGAATAAA TCACAGGCAT TTTATACTTT TTGGCAATAT 780  
 25 CATTTATCGC ATTCATTAAT GATTAAAAAT TCTTTTCATT ATCGATATTC TCTTCTCTAT 840  
 GCGCAGaTAC TAAATGTAT TGTGCGGTT CTAATCCTAG TTTATTTAAA ACGTCACTGT 900  
 GATTAATTTT ATCTCGATGC GCTTCTATCA CTTCTGTCAT CGGTGATCCT GTyACAAAGA 960  
 30 TATTGCTTTT ATTGAAGCCT TCATCTAATA AATAACGTCT GCTATGTTCC GTATAAGGTA 1020  
 GATTACATC ACTGACATGG TCAACAATTT TACGATTGAT TTCTTCAGGT ACATTCTGAT 1080  
 35 CAAAGCATCT ATTACCCGCT TCCATGTGGA ACACAGGAAT CTTTAATCGT TTAGCAGATA 1140  
 CTGCTGCTAA ACAACTATTT GTATCACCAA GAATTAAAAG TGCATCTGGT TGTTGCGGTA 1200  
 ATAAACATC ATATGTCTTC GCAATAATAT TCCCCatCGT TTCTCCAnGt TACTTCCAAC 1260  
 40 TGCCTCTAAG TAGTGGTCCG GTTGTCTTAA TTCCAAATCA TCAAAGAAAA TTTGATTCAA 1320  
 TGTATAATCA TAATTTTGAC CAGTGTGTAC TAATATCTGA TTAnAAATAT TGAT 1374

## (2) INFORMATION FOR SEQ ID NO: 555:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1472 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 555:

55

TTACCTCTAG GCATGTCCCT TTCACGGTTT GCTTATGATA ACGTTATCGA CATTGTAAAA 120  
 TTGTCCTTTT TGAAAAATA ATTCTGGTGA GATAGACGCT ACTTGGaTTT CaTCaGTTTG 180  
 5 TAAtAACGCA GtATAAtTAC CaTTACTAAA tTGaGTTAAT CGTTCaTATA AAGTACTAAT 240  
 AGGATAATAA ATGTTATCCG TTAAGCGCGC CGTATAGTTC ACTTGATACG TTTCGCCTTC 300  
 AACAATTGCT TGTTGGACAC GTTTAATATT AGTCATCATA ACCTCAGAAG ATTCAACAAA 360  
 10 TGAAAAATGA TACTTTGATA CATAAGAAGT TTGATGTTCA TATGTTGAAT TTATGCTTTC 420  
 CGCTTTTTTCA AAActATAAG CTGCTGCATA AATATCATCT TTAGCTAATG AATGTGTACA 480  
 CATAGCATGA TTAAATACT TTGCCGCTTC GTAActTAAA TATAACGAGA CATATCTACC 540  
 15 TTGTCGtKgt GCGCTTGTC AAAGTGATC ACTTCTCCA CATCAGCCAA CTTAGTAGCA 600  
 ACATACTTCT TTATAAATCC CTTTAATTGA ATATGGTATT GCTTATATTC ATTTTCAGTT 660  
 20 AAATAGTAGC GATAATTATA TTCTATTCTC ACAGTAATCA CCTACCTTCG ATAAAAATAA 720  
 TTCAACTTGT CGATAACCGT ATTCACTCAA AATAGATTCA GGATGATATT GCACACCAAA 780  
 AACCGGAAAT CTAATATGCT CAAATGCCAT AATAATCGCT TCATCGTTTT TTGCTGTAAT 840  
 25 CTTTAAGCAA TTTGAAAAG TCGCTCCGTC AGCAATTAAT GAATGATAAC GCATTACATT 900  
 GAAATTTTGA GGCAGTCCTT GAAAAATACC TTCATTGGTA TGGCGTAACT GTGTAGTATG 960  
 TCCGTGTACA GGATGATAGC CGTGAATGAT ATTTCCACCA AAATAAGACA CGATACATTG 1020  
 30 AAATCCTAAA CATAACCTA GTATAGGTAC ACGCTGATAA AATTGTTCTA ACACTTCATT 1080  
 CAAGATAGGA TAATCATCCG GaTTACCCGG CCCAGGCGAA ATAACAATTG CTTTTGGCTT 1140  
 35 CATATTAATG ACGTCTTCTA TCAGCAGATT ATCAATACCA ACAACTTGAA CTGTTAGTTT 1200  
 CGTTTGAGTC TTAATATAGT CTATTAAATT ATATGTAAAT GaATCATTAT TATCTATGAC 1260  
 TAGAATCATT GTATACTCCG TTCTAAATGT GTTTATTTT TATAATATGT ATTGGATGTA 1320  
 40 GCTAAAACCT TAAAAGCATT GTCATTATCC TGACCTTGAT TTAACTAAT ATATGGTATA 1380  
 TTCTATTCAT CGTACATAAA TGAATATCAG AGGTTCCCTAG CTGAAACCCT CTATAAAAAA 1440  
 CTAGGCCATT GAAATTTCAA ACATTGTTG GG 1472

(2) INFORMATION FOR SEQ ID NO: 556:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1054 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

AGAACAGCAA GGATTACTTA CTGAGGAATT AAAGAAGGAT ATTTTAAAAC AGAACAAATT 60  
 ACAACGTGTT GAAGACCTAT ATAGGCCTTT TAAACAAAAG AAAAAGACAA GGGCAACTGA 120  
 5 GGCGAAACGT AAAGGGTTAG AGCCATTAGC GATATGGATG AAGGCACGTA AACATGAAGT 180  
 CTCAATTGAA GAAAAAGCAC AACAATTTAT AAATGAAGAA GTGCAATCGG TTGAAGATGC 240  
 TATCAAAGGT GCACAAGATA TTATTGCGGA ACAAATTTCA GATAATCCTA AATATAGAAC 300  
 10 AAAAAATTTA AAAGATATGT ATCATCAAGG TGTGTAACT ACATCTAAA AGAAAAATGC 360  
 TGAAGATGAA AAAGGTATTT TTGAAATGTA CTATGCATAT AGTGAGCCAA TTAAACGCAT 420  
 15 TGCTAATCAT AGAGTTT TAG CTGTTAATCG TGGTGAAAA GAGAAAGTAT TATCTGTAAA 480  
 GTTTGAATTC GATACGACAT CAGTAGAGGA TTTCATTGCA CGTCAAGAAA TCAATCATAA 540  
 TAATGTAAAT CGCAGTTATA TTTTAGAGGC GATTAAAGAT AGCTTGAAAC GCTTAATTGT 600  
 20 CCCTTCGATA GAGCGTGAAA TCCATGCTGA TTTAACTGAA AAAGCTGAAA ATCATGCAAT 660  
 AGATGTTTTT AGTGAAAAC TAAGAAATCT ATTACTGCAA CCTCCAATGA AAGGTAAACA 720  
 AATATTAGGC GTAGATCCAG CATTTAGAAC AGGTTGTAAA TTAGCAGTCA TTAACCCATT 780  
 25 CGGTACTTTT ATAGCAAAAG GTGTGATTTA TCCGCATCCA CCAGTTTCTA AAAAAGAGGC 840  
 AGCAGAGAAG GATTTTGTAC AAATGGTTAA AGCGTATGAT GTGCAATTAA TTGCAATTGG 900  
 CAATGGTACT GCAAGTCGTG AACAGAACA ATTTGTTGCA GATTTAATTA AAAAGCATCA 960  
 30 GTTGCCAGTA CAATTCATCA TTGTCAATGA AGCGGGCGCT TCAGTATACT CAGCATCAGA 1020  
 AATTGCTAGA GATGAATTC CTGATTTTCA AGTG 1054

35 (2) INFORMATION FOR SEQ ID NO: 557:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1057 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 557:

45 AAATGTCAGA ATACAAGAAA AAAATAATTG AATTAATTGA AAGTAATTTA ACAGGATATG 60  
 AAATTTCTAA AAAAAGTGA GTTTCTCAAT ACGTACTTTC ACAATTAAGA CAGGGCAAAC 120  
 GCGAAGTAGA TAATCTAACC CTGAATACAA CAGAAAAATT ATATGAATAT GCCAATAAAG 180  
 50 TTTTGTAATT TAACTAATGT ATAAATTAAT CAAGCTATGT TTATTTGATT TAACTATTAA 240  
 TAAAAATCAT ATGGTGAATG GATATTATAA TAATTAAAT ACAAAAATAG TAGATTCCAA 300

55

AAGGGAAAAT AAGTGTTAAG TTTTAAATGA TAAAAAAGAT TGGAATGGAT CGTCTTGAAA 420  
 TGCTCCCTTC AAAGTTTTCA TTTTTTCAAT GTCGACTTCG AAGGGGGCAT TTTCATTAAA 480  
 5 TTGTTATAGC TTTTATATT TGTATAATGA ACATATAAGT TTAAGAAGGT GCGAGTGAAG 540  
 GAAATAAAAA AGCTCAAATG TACCAAATTG TTAATCTTAA TAAATCTCTA CTTTATAAAG 600  
 ATTGAATGGA CATTGAGCG TTAATCAGTC AGGAGGGACT TTCCCTCCTA CAATTTAATA 660  
 10 ATAATACTTG CTTCACTACT ATACAAGGAG TGAGTTGTTA TGTTCAAAGT GAATTATTCG 720  
 ATTTTAAGTT ATTATCCAGA ATATAATATC GCAGTAAGTT GGCAACGTTT AAGAGAAGGA 780  
 15 AAAACAATAA AAAACAAGAT TTAATACTGC TGCCTCATGA GCGCTTGAA CATTATTTGA 840  
 TGAATAAGTA TAATTTCAAC TATGATTATG CACATAAAAT TGTATCAAAA AAATACGATT 900  
 ATTCAATTTT TATAAAAAG AAGGTGGATT AAATGCTTAC ATTAATAAAA TTGGAAAGAA 960  
 20 GATGaACAGG tTATAATATA TGraTATATa CCTGaAGATG aTATAAGTAC CGGGTAAAGG 1020  
 GTCCCGTACC TTTTAAATTA AAAAGTTCC AGGGGGT 1057

(2) INFORMATION FOR SEQ ID NO: 558:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3754 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 558:

35 CTGTGCTGTA TTTACTTTAA TTTGACGACA TTTGAGAAGT ATTATGATGG nTGAGTTAT 60  
 TTTATGAAAG TAATGTATTA ACAATCGAAT TACTAAGTCT ATATTGGTAG GAATATCGAT 120  
 GTTTAGTTTA AATGGAATGC ACTATTTAAG TTTTAAATAT GGAGATGTTT GTGACTTTTC 180  
 40 GATGATTAAAG ATTTTATAG GTGTGCATCA TTTCAAATAA ACTTTGTGTT TAAAATTGAG 240  
 CTTAGGAAAT CGATAGGTTT AGATGAGGAT ATTGTTGAAG TTATGTGTCT TGTATCCTTA 300  
 GTTGTATATAA AAGCGCAAAA AATAGCACCG CTTTCTCTTT ATCTGTGTAG AAAGGATGCT 360  
 45 ATTATTGTAA AACAATAGTT TTAATTTnAT TTTCTGATaT ATCATATGTt ATTcTACCTG 420  
 TATCAATTTT TATCGAATTA TAACCATCAA AATTATCAAC TTTATCATCA AAGTCTATCA 480  
 CTTTCCAAT TAATATTTTA TTATTAGTAA GCGTTAATTT GACTAATTG CCTATGTATG 540  
 50 ACTGTAAATT CATATTTAAT CACTCCTTTT TAATATACGG AACTACATGA AAACCAGTTT 600  
 TAGAATAATG AACCTTACCT AATTCGTTT CAATATATTT ACCATTCACA TAAGATTIAC 660

55

	AAACTGGAAT ATCAATAAAAT TCTTTTITAG TCATCTTTTC ACTTTCATTA GAATCTATTA	780
	TAGTATAACT TGGTAATAAA GATGAATTAT TTTTCTTATT ATTATATTCA TTTGTGCATA	840
5	AAATATGACG CTTTTGCTTA ACATTATTCA ATTTTCATCGT CATTTTACCA TTACTTATCA	900
	TTTCAAGCAT TTCTTTTITTA GCTTTTGTGA GAAAGGATGC TATTTTGTAA TTATTTTAAT	960
	AATCTAATAC TTTTATCTC AGTTTCATCA AACGAATAAA CAGCAAAATC AGTCTGTATA	1020
10	TCTACAACAA GATTCCTGT TTCGCTCTCG AATTCATTTT CATAGTCAGT TACAAACCCT	1080
	TCGTATTCTT CATTATTCAA AAGCGTGATA ATAACATCTT TACGATATGC ATCTnCAATT	1140
15	CTCAATACTT TTCACCTACT TATCAATATA AGGTACTATA TGaGcACCTG GTCTTCGAAT	1200
	AATGGCACTT cCCTTTCTGG TTTCAATATA CACATTTTCG ATATGTATTT TTCCAATAAT	1260
	TTGATTAAAA TTAATAATCT CTTTCAAATC AAATCGCTCA TCACTTAATA TTAGATTGCC	1320
20	TGTTGACATT TTTTCTCTTA ACAATTCATT CAATAAATCT ATAGAAAGTA TTGTATAGCT	1380
	AGGCAATTTC TTATTATTTA AAATGGCTCT TTTTTTATTT TCATTATATA GGTGATGACC	1440
	TAACATATAT CTATTTTATT TTCCGCAATT TATTTCTATT TTTATTTTAC CATTTTTAAT	1500
25	CATTTCTTTC ATCTGGTTTT TAGCTTTTTc CTGtAATTAT GCTTCTTTTA CTTCTACTTG	1560
	ATATTTACct TCACGCTCTT TAAAGAACTT GTCCCGCCAA TTGCCAACAT GTGGCACTGT	1620
	GGTACTTCTA CACCAAGGAT GCATAGGTGG CGCATTCACA CCTGGTATCA TATCTTTAAC	1680
30	TTTAAATATT TTTCCGTTAA GTGAATGACA TAATTTAGAT GTTTTACTAT CTATTTTGGC	1740
	AACATATTTA TATTCGCCAT CTTACCAAG TTCTTTTAAA TATGTTAACT TTTGTGCTTC	1800
35	TGCATTTTCA GTAAATAGTT AAAAAAGCGT ATAAAAATAG CACCACTTTC TCTTTAkCTG	1860
	TCTAAAAAGG ATGCTATTTA TCTTTTGAAT TTGAATCTT TTTGCTTTT TCTATACTTT	1920
	CAAATTCCTC AACTAATTCT TTAAGAGATT CACTCAATTC TTTTGCAGTT ACATTTCCAT	1980
40	CTAATTGTGA ATCTAACATA ATTAAAATCA TCTCACTTTA TATTTAATCa TATTTATACT	2040
	ATAAGTTTTT TTCAATAATT TTTCAATATG GCTATCATTA ATGATATTAA TATGCGTAAA	2100
	ATATTTAGCA CAAAATTTAC TCACTATTTT ACCATGAAAC CTATTTGACT TGGTAATAAA	2160
45	TTTTACTTGT CCCTTATTAG TAACGATTGT CATTGATTTT ATTGATGGAT GCTTAAAAAA	2220
	TGTAAATAAA TCATATTCTG AAAATCCTGA CTGTCCAGGA TGGTTATGTA ACATAACAAT	2280
50	TGAATTCGGT TTAATGTTAA ATAATAATTC GGTGCTTGT TACCCTGGCA CAAAAGATAC	2340
	ACTATCTTGA TTGACATATA CTTTTGTAAA TTTACCATCT TTTAACAAAT AAGCTACTTC	2400
	ATTGCTATCA TTGTTTTCCA TGAGATATAC CTCCGTTTAT AGtCTGTGGA CTGATATTCC	2460
55		



GAAGGTAAA ATTTTAAAG TAACTTTCT TTATCCCAT GCTACGAGTT CAGATTCAGG 2580  
 AAATAGCCCT TTACTAGTAT TTATGTATAT TCTGTCTATG GCATGAATAA AATAATTATC 2640  
 5 TCTTGATTTT tTTTCTAAAC TAGATTTTTC AGCATTGATA ACTTCAAGAC TATCTATATC 2700  
 CATTGAATA ATACCAGGCT TAATATTTTC ATCATTATTA GGAAAATATT TATATGTAAC 2760  
 ACTTTTATCA TTAATTTCTT TTATTTTAA TATTAGCAAT CATTTCCACC TCTAATTAAT 2820  
 10 TAAAATACTA TAATTATATT TTATTTCTGT AAGTTTATGT GCCTCTATAT AGTGTAATTT 2880  
 ATACTTATTC ATTAGATAGT GTTCAAGAGC TTCATGTTTC TACATTATTA TATCCATTTT 2940  
 15 TTTAATATTT TTCCCTTCTC TTAAACGTTG CCAACTTTGA GCCATATAAA AGTCAGGATC 3000  
 AAATGTGTTA AATCCACTTT CTAATAAATA CTTATTTTCA AATATATGTT CATAAACTCT 3060  
 TTGAATTA AAA TTTTATTTA TATTAgTATT TTTAGCAATT TTAGAAATCT CTATCTGTTT 3120  
 20 ATCTCGATTT CTAAGTGAAT TATAATAAAT TTGAGCATGT CTGTTCCCTT TGATACCGTA 3180  
 TTCATCACTT TTATTATTAA GTGCACCTGA TTCAATAAAA CAACCTTCTA CTTGATATTT 3240  
 ACCTTCACGC TCTTTAAAGA ACTTGTCTCG CCAATTGCCG ACATGTGGCA CTGTGGTACT 3300  
 25 TCTACACCAA GGATGCATAG GTGGCGCATT CACACCTGGT ATCATATCTT TAACTTTAAA 3360  
 TATTTTCCG TTGAGTGAAT GACATAATTT AGATGTTTTA CTATCTATTT TGGCAACATA 3420  
 TTTATATTCG CCATCTTCCA CCAAGTTCyT TTAAATATGT TAACTTTTGT GACTCtTCyT 3480  
 30 TTTCAACGAA TAATGAAAAA AGCATATAAA AATAGCATCG CTTTCTCTTT ATCTGTGTAG 3540  
 AAAGGATGCT TTAATACCaT GCTATTTTAT AATTTTcGGG AAATTcTTGC TTCTCGATAA 3600  
 35 AGTCTCTTAC TACAGAAAAA GACTTATTAC GATATAACAT AAAATATTCT TCATTTTCTA 3660  
 TTTAGATAA TGAGAAATCT ATAACCTCTG CATCTTTTTT ATTAAAAGTT ACTGAACCTT 3720  
 TACCGTTACT TATATCATCT TCAGGTATAT ATTC 3754  
 40 (2) INFORMATION FOR SEQ ID NO: 559:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 815 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 559:

50 ATTTAACTAA AACTATAaAT AATCAAATGA TATTGGAAGA TATTaGCATA GATATCGAAA 60  
 AAGGTAAATT GACTTCTTTA ATTGGACCTA ATGGTGCGGG TAAGAGTACT TTACTTTTCAG 120

55

CTGATTATAA AAATAATGAC TTGTCGAAAA AAATATCTAT ATTAAAACAA ACAAACCATA 240  
 CTGAAATGAA TATTACGGTA GAGCAGTTGG TAAACTTTGG ACGATTCCCT TATTCTAAAG 300  
 5 GTCGTTTGAC GAAAGAGGAT CATGATATTG TCAATGATGC GCTAGATTTG TTGCAACTAC 360  
 AAGATATCAG AAATCGTAAT ATTAAGTCAT TATCTGGTGG ACAACGTCAG CGTGCATACA 420  
 TTGCAATGAC AATAGCACAA GATACTGAAT ATATTTTGCT AGATGAACCA TTAAATAATT 480  
 10 TAGATATGAA GCATGCTGTT CAAATTATGC AAACGTTAAA AATGTTAGCG CATAAAATGA 540  
 ATAAAGCGAT TGTCATTGTG TTACATGATA TTAACTTTGC GTCCTGTTAT TCAGATCAGA 600  
 TTGTAGCATT GAAAAACGGA CAACTAGTTA AGTCAGATTT GAAAGATAAT GTCATTCAAA 660  
 15 GTAGTGTTTT AAGTGATTTA TATGACATGA ATATTCAAAT TGAACATATA AGAAATCAAA 720  
 GGATTTGTTT ATATTTTAAG GATTGATAAT TTGGAGaCAC TTTAAAGGGG TGATGCGCCA 780  
 20 ATTAAAGAAG GGTAAACGT AAAGCATTTA TTTAT 815

(2) INFORMATION FOR SEQ ID NO: 560:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 919 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 560:

GAAACGAATA ATAAATTTAC TGAGTTATTA GTTGAAAAAG CTAATAAACA TGATGATGTT 60  
 35 CTCGATAmGT TGATTAAATAT TTTAAAATAA GCGATACACA CTAATAAAAT TGTATTATTA 120  
 TTATGTTnAA TTGACnCCCTC CTAAATTTGC AAAGATAGCA ATTTAGGAGG CGTGTTTATT 180  
 TTTATTGACG TCTAACTCTA AAAGATATAA ATTAGACATT TACAAATGAT GTAAATAACG 240  
 40 CAATTTCTAT CATCGCTGAT AACAATTCAT GGTTTAATAT GCAATGAGCA TATACTTTTT 300  
 AAATAGTATT ATTTACTAGT TTTAACAATC AATTAATTGG TATATGATAC TTTTATTGGT 360  
 TATTTTTATC CCATAGTGTG ATAATTACTA TTTTTCATTC ATAATAAAGG TTTAAAGCAT 420  
 45 GTTAATAGTG TGTAAGATTA ACATGTACTG AAAAACATGT TTAACATAAT GATATAAGGA 480  
 GTGACGTACA TGATCCGTCT AGGTAAAAATG TCAGATTTAG ATCAAATCTT AAATCTAGTA 540  
 GAAGAAGCAA AAGAATTAAT GAAAGAACAC GACAACGAGC AATGGGACGA TCAGTACCCA 600  
 50 CTTTTAGAAC ATTTTGAAGA AGATATTGCT AAAGATTATT TGTACGTATT AGAGGAAAAAT 660  
 GACAAAATTT ATGGCTTTTAT TGTTGTCGAC CAAGACCAAG CAGAAATGGTA yGnyGACATT 720

55

TATAAAGGAG CTGCTACAGA ATTATTCAAT TATGTTATTG ATGTAGTTAA AGCACGTGGT 840  
 GCAGAAGTTA TTTTAACGGA CACCTTTGCG TTAAACAAAC CTGCACAAGG TTTATTTGCC 900  
 5 AAATTTGGAT TTCATAAGG 919

## (2) INFORMATION FOR SEQ ID NO: 561:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 518 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 561:

ATCATATAAA CCGGCTGCTT CTAAAAACGA AAATACTGTT ACTGGACCTA AAAATTTAAA 60  
 20 CCCGTATTGT TTTAAATCTT TAGATAGTTG TGTTCAGTA TCATCAACTG TGATACGATC 120  
 AGAAGCATGT TCATACTGCA AATCTTTAGG CTTACCATTT ACATATGACC ATAAAAATTT 180  
 ACTAAACTA CCATATGCTT GTTCAATTTT TAAATACCCT TGAGCTTGAT TAACAATTGC 240  
 25 TTCTAATTTT TTACGATGAT GAACGATATT TGGAAAAGTC ATTAAGCGGT CGATATCTTG 300  
 AGCGGTCATT TGTGCTACCT TTTCTGGTTC GAAATCATAA AATGCTTCTT CATAGGCTTC 360  
 TTTCTTTTTT AAAATAGTTA ACCAAGATAG CCCAGCATGT TGTGATTCTA ATGCTAAAAG 420  
 30 TTTAAACAAT GCCTTGCAAT CATAGAGCGG TTGTCCCAT ACATGGATCm TGATAGTCTA 480  
 AGTAGACTGG GATCTTTAGT ACCAAATGCG CATTCAAT 518

## (2) INFORMATION FOR SEQ ID NO: 562:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1539 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 562:

CTTATTTAAA AGTAAATCAA TCAATGTATT ATAATCCGAA TAGTCCGCAT AAAGCTGGTT 60  
 TGCaGGCAAA tCAATTACTA CAACAAGCAA AAACCCAAAT TAATGCAATG rTTaATTCAA 120  
 AAACAAATTA IGATGTTGTA TTCACTAGTG GTGCActGAA TCCAATAATC TTGCTTTAAA 180  
 50 AGGTATTGCC TATCGTAAAT TTGATACAGC GAAGGAAATA ATTACATCCG TGTTAGAGCA 240  
 TCCGTCCGTA TTAGAGGTTG TAAGATATTT GGAAGCACAC GAAGGATTTA AAGTTAAATA 300

CAAAGTCGGT TTAGTAACAT GTATGTATGT AAATAATGTA ACTGGACAAA TACAGCCTAT 420  
 TCCACAAATG GCTAAAGTTA TAAAAAATTA TCCTAAGGCA CATTTTCATG TAGATGCGGT 480  
 5 TCAAGCATTG GGCAAAATTT CAATGGATCT CAATAACATA GATAGTATTA GTTTAAGTGG 540  
 ACACAAGTTT AATGGTTTAA AAGGACAAGG CGTCTTACTT GTAAATCACA TTCAAAATGT 600  
 TGAACCAACT GTCCATGGTG GTGGTCAAGA ATATGGTGTT AGAAGTGGA CAGTTAATTT 660  
 10 GCCAAATGAT ATTGCAATGG TTAAAGCGAT GAAGATAGCT AATGAAAAT TTGAAGCATT 720  
 GAATGCATTT GTTACTGAGT TAAATAATGA CGTCCGTCAA TTTTAAATA AATATCATGG 780  
 15 AGTTTATATT AATTCTTCAA CTTCAAGTTC ACCATTCGTT TTAAATATTA GTTTCTCTGG 840  
 CGTAAAAGGT GAAGTATTAG TTAATGCTTT TTCAAAATAT GACATTATGA TATCTACGAC 900  
 AAGTGCTTGT TCATCTAAAC GTAATAAATT AAATGAAGTA TTGGCTGCAA TGGGATTATC 960  
 20 AGACAAATCT ATTGAAGGTA GTATAAGATT ATCATTGCGG GCTACTACAA CTAAAGAAGA 1020  
 TATAGCGAGG TTAAAGAAA TATTTATCAT CATTTATGAG GAAATTAAGG AGTTGCTAAA 1080  
 ATAATGAAGT ATGATCACTT GCTTGTTAGA TACGGGGAGT TAACATTAAA GGGTTCAAAT 1140  
 25 AGAAAGAAAT TTGTAAATCA ATTAAGAAAT AATGTAAATA AGTCaTTAAA AGGACTTGAT 1200  
 GGGTTTGTCTG TTAAAGGCaA ACGAGATCGT ATGTATATTG AACTTGAAGA CCATGCaGAT 1260  
 ATAAATGAAA TAACATATCG ATTATCAAAA ATTTTCGGTA TTAAATCTAT TAGTCCAGTA 1320  
 30 TTAAAGTAG AAAAAACAAT AGAGGCAATA AGTGCAGCGG CAATTAAATT gCGCAGaATT 1380  
 TGAAGaAAAC AGCACATTTA AAATTGATGT GAAGCGTGCC CGATTAAAAT TTCCCAATG 1440  
 35 GATACGGTAT GGAATTACAG CGTGGAATTG GGGTGGTGCC AGTATTGGAG CACTTCGCCA 1500  
 TATTTCCAGT GGATGTCCAA CGTCCCAGnC CCAGGAATT 1539

## (2) INFORMATION FOR SEQ ID NO: 563:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 968 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 563:

50 ATAnCGTATA CATGTGTTCT TTAAATTGT GATAAGGAGT TTAGGATGGT TTATTTAAAA 60  
 TCAATAGATG CCATTGGATT TAAGTCTTTT GCAGATCAAA CCAATGTTCA ATTCGATAAA 120  
 GGTGTAAC TG CAATTGTTGG TCCAAATGGA AGCGGTAAAA GTAATATTAC AGATGCTATT 180

55

ATCTTCTCAG GTGCAGAACA TCGCAAAGCT CAAAATTATG CTGAAGTACA GTTAAGATTA 300  
 GATAATCATT CTAAAAAGCT CAGTGTGAT GAAAACGAAG TTATTGTAAC AAGAAGATTG 360  
 5 TATCGAAGTG GTGAAAGTGA GTACTACATA AATAATGACC GTGCAAGATT AAAAGATATT 420  
 GCCGATTTAT TTTTAGATTG TGGATTGGGA AAAGAAGCGT ATAGCATTaT CTCGCAAGGT 480  
 AGAGTTGATG AAATACTAAA TGCTAAACCA ATTGATAGAC GTCAAATTAT TGAAGAATCG 540  
 10 GCTGGTGTAC TTAAATATAA AAAACGTAAG GCTGAATCAT TAAATAAACT TGACCAAACA 600  
 GAAGATAATT TAACGAGAGT AGAAGACATT TTATATGATT TGGAAAGGTCG CGTAGAACCT 660  
 CTAAAAGAGG AGGCAGCTAT AGCTAAAGAA TATAAGACAC TTTCACATCA AATGAAACAT 720  
 15 AGTGACATTG TAGTTACAGT sCAGCATATT GATCAATATA CAAATGACAA TAGACAATTA 780  
 GATCAACGTT TAAATGATTT ACAAGGCCAA CAAGCAAATA AAGAAGCTGA CAAGCAACGT 840  
 20 TTAAGCCAAC AAATTCAACA ATATAAAGGT AAACGTCATC AACTTGATAA TGATGTTGAA 900  
 TCgCTTAATT ATCAATTAGT AAAAGCTACG GAAGCCTTTG AAAAATATAC GGGACAATTA 960  
 AATGTTTT 968

25 (2) INFORMATION FOR SEQ ID NO: 564:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 436 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 564:

TTGTGTGTAT GCATTCAATG TGCTCTGTTT GTAAATGGCT AGCTATATAA TTTAgGATTC 60  
 GAGGATCGTC ATCGACAACA AGACATTGCA CCATAGCTAT AAACTCCCTT ATCTTTTTCA 120  
 40 TTTATTATAC ATGTAAAATA TTTTTCGTA AAAAAACAAT TGTTCAATATT GAGTTCATAT 180  
 TTCAACCTTA TACTGACGCT AAAGAAGAAA TAGGGAGAAG TGAATCGATA TGAAATTAGC 240  
 GATAAAAGAG ATTATGTTTT ACAAATTCG TTATATTTTA ATCACATTAA TCATTCTTTT 300  
 45 ATTAAGTATT ATGGTGTTAT TTATTAGTGG TTTAGCTCAn GGGCTTGGTA GGGaGAATAT 360  
 TTCGTwATTT GAACACTTTG GATAATGATG aaTATGtTGT TCaAAAAATG AAAGAGCCGC 420  
 aAATTGaGAA ATCGCA 436

50 (2) INFORMATION FOR SEQ ID NO: 565:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2554 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 565:

	ATGTTTGTGTC AATATCAATT GTTTGAGATA AATCCGCTTG TATAAAACGA ACTTTATCAT	60
10	CATTAAATTT GCTAGTTAAT TCATTTATAT CAGTACGATA ATATTGCACA TAAACTTCAA	120
	ATCCATCAGT TAATAATTGT TTGACTATCT CAGAACCAAT TGAACCAGAA CCACCTAATA	180
	CTAATGCTTT CATTACTTTT TAATCTCCAA ACGACTATCG ACTTGCTGAT CTAAATTTAA	240
15	ATATAACGAT GACGTTTCGT TAATACTATC TAACGTGATA TTTTCAACAA TGTCTAACAT	300
	GTCAAACACG CTAACACCTT CAAAATACAA TTAGTATAT TGATTAGCAA TATATTGAGG	360
	TGAGTTTAAA CTTGATATGA ATTCACCTAT AAATTGCTTT TTCAAAAGTT CAAATGCTTC	420
20	TGCATCTTGG AAATGCGCTT TTTTATCAGC CAACTCATCT AATAATAATT TTTTAAATTT	480
	ATCTGGTTCT TCAGTAGCAC TTGTCACGAT TGAAAAACTA TACGTCGGCT CTAGTACAAA	540
25	TTGATAACCA AATGTATCAT CGATAAGTCC TTCGTTTAAT AAATTCTGAT AAAAATCTGT	600
	TTCTTCCCA AAAATTAAC CAAAGAATAA TGACATTTCT AAATCACGTT GTACATATTT	660
	TTGAGGCGCT TCTTGTAATG GTTTATTTTT AAAACCAAGC ATTAGTCTTG GTGATTGAAT	720
30	TTTCATAGAT TCAGTAACAA ATGCTTCTTT AACATCCTCC GGTTCATCAA CAAGTCCTCG	780
	TTCGATTTTG GGTGGTTAA CTTTATTACG AGCATCCTCG TGTGTTTTA CTATTCgACA	840
	TATTGCTTCA GGATCCACAT CGCCAACAAC AAATAAAACC ATATTTGATG GATGATAAAA	900
35	CGTTTCATAA CATAGATACA AATCATCTTT TGTAATATCG TATATACTTT CTACACTACC	960
	GGCAATATCA ACACGTATTG GATGTTGTTG ATACATTGCA CGCAATGTAT TaAACATTAA	1020
	TTTATATCCA GGTGCTCTT GaTACATTTT TATTCTTCT GCAATAATAC CTTTTCTTT	1080
40	ATCAACAGTT TCTTTTGTA AATAAGGCGT TTCaACCATT GTAAGTAAAC GTTTAATGTT	1140
	GTTTTCAATA TtATCAGTTG CACTGAACAA GTAGCTTGTA CGATCAAAGC TTGtAAACGC	1200
45	ATTTGCTTGT GCGTTATCTT CAGCAAACGC AGTAAATAAG cTTCTTCTTC TTTTTCAAAT	1260
	AATTTATGTT CcTAAAAAGT GAGCAACTCC ATCAGGTACA GTAACAAATT GGTCTTGTC	1320
	AAGGGGTTTG AATTGATTAT CTAATGAACC AAATTGTGTA GTGTAAGTGA CAAATGTCTT	1380
50	TTGAAAACCT GGTTkGGGGA TAATAAATAA TCGTAAACCA TTTTCTAATt CTTGTTCGAA	1440
	tACTCTTTTCG TCTATTAATT CATAATAACG CTCTTTCATT ATTTATCCCC TCCTTTTGTC	1500
	AAACATAGA TtGTATCTAA AAATGCTTTT TCAGCAACAG AAACAATATC TTCGCGACTT	1560

55

TTATGCATAA TCTCTATAAT ACTTTTCGGA CGATCTTCAG ATTCATATCG ATGAGAAATG 1680  
 ATTACTTTTT TAGCTAACTC TAATTTTTCT TCAGTGAAAT CTCCTGCTTT TATTTTTTCA 1740  
 5 AATCACTTA TAATAGTGTG TTTTGCAGTT TCGTACTTAT CACTTGAAAC CCCACTCAAA 1800  
 ACAATAAAT AGCCATTTTT GCCATCAATT TGTGAATGTA TAGAGTACGC TAAACTTTGC 1860  
 TTTTCTCGCA CTTCAATTAAA TAAAACAGAT GAAGGATCTC CTCCAAACAT CATGTTAAAT 1920  
 10 ACAACAAAGG CAGCATATCC ACTTTGTCCA TATTGTGTTG GAAAACGGTA TCCCATATTT 1980  
 AATTTAGCTT GATCCACGTC ATCATATTCA ACAATATAAT CAACTTCTTC ATCGTGTAAT 2040  
 15 TGATGAGTAG AATGTTGGAA TTGATGTTTA TCGAATGGTT TAAGTGCAA TTTTTCACGT 2100  
 ATTTGTTTCT CAACACTTTC AGGTTCTACA TTGCCGACAA CATAAACAGA ACATTGATCA 2160  
 TTATTAATCA TTGATTGATA TGTATGATAT AGTGTTCAG CAGTAATATG TGGGATTTGT 2220  
 20 TCTAGTTGTC CTGTAGATAA GTATTTATAT GCTTCATTTT CAAACATATG GTCGAGTAAT 2280  
 TTTAAAACG AATATTGTGC TTTATTATCT ACCATTGCTT CTATTTTTTT GGCTAATAAT 2340  
 GTTTTCTCTT GGTTAACAAA ATTATCATTG AATGCTTTAT TTTCAATTAA TGGATTCCAA 2400  
 25 ATGATTTCTT GtAATAAATC TAATCCTTGa TTAAATAATG AwTCACCGkT TcyTAAATAA 2460  
 CGkkCaTTAA caATTyCTAA tGaAAATGtA ATgACaTGCT GaTCTTTGAA TTTTGAAATT 2520  
 30 GTACTATTCA CATACGCACC ATATAAATCG GCTA 2554

## (2) INFORMATION FOR SEQ ID NO: 566:

## (i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 1424 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 566:

40 TnTCGCTATT TtnAATTGGT TTTGTATGGT TTAAGTTATA TCAATATACA ACAnACCCTA 60  
 AAGCTGATAT CCCAGGTATC ATTTTtagTA CGATTGGTTT TGGTGCTTTG TTATATGGTT 120  
 45 TCTCAGAAGC TGGCAACAAA GGTGGGGTT CAGTAGAGAT AGAAACAATG TTTGCGATTG 180  
 GTATTATCTT TATTATTCTA TTCGTTATTA GAGAATTAAG AATGAAATCA CCAATGTTGA 240  
 50 ATTTAGAAGT ATTGAAATTC CCaACATTTA CATTAAACAAC AATTATTAAT ATGGTTGTAA 300  
 TGTTAAGTTT ATATGGTGGT ATGATTTTAT TACCGATTTA TTTACAAAAT TTACGCGGAT 360  
 TcTCAGCATT AGATTCCGGA TTGTTATTAT TACCTGGTTC TCTAATTATG GGTCTACTAG 420  
 55

TTGCTGTAAT GACTTATGCA ACATGGGAAT TAACTAAATT AAATATGGAT ACACCATATA 540  
 TGACAATCAT GGGTATCTAT GTACTTCGTT CATTTGGTAT GGCATTTATA ATGATGCCAA 600  
 5 TGGTAACTGC AGCTATTAAT GCGTTACCGG GACGACTTGC CTCTCATGGT AATGCTTTCT 660  
 TAAATACGAT GCGTCAATTA GCAGGCTCTA TAGGTACAGC AATCTTAGTT ACTGTAATGA 720  
 CAACACAAAC TACACAACAC TTATCAGCTT TTGGGGAAGA GTTAGATAAA ACGAATCCTG 780  
 10 TTGTACAAGA TCATATGCGT GAATTAGCAT CACAATATGG CGGACAAGAA GGCGCAATGA 840  
 AAGTGTACT ACAATTTGTA AATAAACTAG CAACGGTTGA AGGTATTAAT GATGCATTTA 900  
 15 TAGTTGCAAC GATATTTAGC ATCATCGCCT TAATTTTATG TTTATTTTAA CAAAGTAATA 960  
 AAAAAGCAAA AGCTACAGCT CAAAAGTTAG ATGCAGATAA TAGTATCAAT CATGAATAAA 1020  
 TAAAATAAAT TAATTGAAGT GTGACTAATC AAAAATTATG TTGTGGGGAC ATGATTTTAA 1080  
 20 AAGTATCGGT GCCAAATATG GTTATCGATA CTTTTTTTAT TTGTTGATTT ATAGAATGTT 1140  
 AGAGGAATTA TATTAAATT TGGCATTGAC GTAGTAGGTC ATTAATAAAG AAAAAGCAGG 1200  
 AAGTGGGTCA ACGAAATGAA TTTTGTGAAA ATAACATTTT TGTCCCAATC CCTACTATAT 1260  
 25 AACATTATTT TAAACGAGGC ATGCGATTAC GGAAGAATAA GCTTATAACA AGTAAACCGA 1320  
 TGCTACAGCC AAGTAAAATG ATGCCGTTAT GAATAGCGTC ACTTgCTGTA ATCACTTGAT 1380  
 30 CTGGTGGTAC ATTTAAATAA TATTTTTTGA AAACATCTGC AATT 1424

## (2) INFORMATION FOR SEQ ID NO: 567:

## (i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 676 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 567:

TCTCCCATGT TCTGCTAAAT GACGCATCAC TTTACTTCA TGAGGCGTCA ATACACGTCC 60  
 45 TTCACCAGCA TTCAAACCGA CAACATTTAA AGGCCCATAT TCAATACGAG ACAGTTTCGT 120  
 CACTTGATGA CCAAAATGTT CGAACATTCT TCTGACTTGG CGATTACGAC CTTCTGTAAT 180  
 TGTAAATTTCA ACCAATGTTG TGTTTTTATC TTTATCTTGT TTCTTAACTT TCACTTCAGC 240  
 50 CGGTTGCGTC ATACCATCTT CTAATTCAAT ACCTTTTCT AGCGCTTTCA CTTCTTCTCT 300  
 CATTAAATAA CCTTTTAATT TCGCAACATA TTTTCTTAA ATTTGATATC TTGGATGTGT 360  
 CATTAAATTA GTAAATTCAC CATCATTTGT GAGTAATAAC AATCCAGAAG TATCATAGTC 420

55



ACGTCCTCTA TCATCAGATA CACTTGTKAT CACTTGAGTT GGkTTATGGA AkAAAAATGka 540  
 AAtTTTGTCT TCTAGTTCTA TTtTAATACC TTCAACTTCA ATCGTATCTG ATGGCTTCAC 600  
 5 TTTTGTTCCT AATTcAGTGA CAGTCGTACC ATTCACTTTC ACTTTTCCTT CAGAAATTAA 660  
 AGTTTCTGCC TTACGT 676

(2) INFORMATION FOR SEQ ID NO: 568:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 454 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 568:

20 GAAACGGTTC TACCAAAAAA CAGTAAGGGC TAAACCCAAT CATGGTAAGA CAAAAAGTAC 60  
 AAATAGCCAT GCCCAAGTTG AACTCGCTGT ACGCCTATTT CTTTCTAAAA AGATAATAAT 120  
 AAAAGCCAAT ACTAAATTAA TGATGAATCC AATGGCTAAA ATAATAGTAA ATAACGTTCC 180  
 25 TAAATCGTTT GAAAATGTAA ATCGCATAGT CTTTTCTCCT ATAAAGAAAG GCACAAAAAA 240  
 ACATTTTGCA CCTTTCACGT CATATTATTT ATTcACAGAT AAAGTTAAAA TTGcATTGAA 300  
 TTCTTCTTCA TTATTTGGGA ATGTTCTTTC TTCTATTTCT TTAATAGTAA TATTTACTAA 360  
 30 TTTTAAATTT GTAGCTTCCT CAGAACTTAA AAAAGCATTa ATGTTTTTTT CTAATAACTC 420  
 kAAAGTCTCA GCTGtAAAaG TTTTAAGTTT AATT 454

(2) INFORMATION FOR SEQ ID NO: 569:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 894 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 569:

45 ACGATATTAC CACCCTCTGA AATAACATCA ATCATGTGAT CTGTAAAGTC CCAAGGATGA 60  
 CTTGTTGTGA AACGAActCT TGAATCGCT ATTTTAGAAA TTGCTTGTAa AAGATCTCCT 120  
 50 AAGTCATATT CTATATCCTG TAAATCTTTA CCATAAGAAT TTACATTTTG ACCTAAAAGC 180  
 GTTATTTCTT TGTAACCTTC ACGAGCAAGT TCACGTACTT CATCTATAAT GTCTTCAGGT 240  
 CTACGGCTTC GTTCTTTACC TCTTGTAAT GGAACAATAC AATATGTACA AAActTATCA 300

TCAATAACGT CTCCTTCTTT AGACCATACT TCAACAACCA TTGCTTTAGA TAAGTATGCT 420  
 TCTTCTAAAA TTTCTGGTAA ATGATGAATA TTATGTGTAC CAAATATCAT ATCTACATTT 480  
 5 TGATACGATT TTAAAATTTT ATTCACTACT GACTCTTCTT GTGACATACA ACCACAAACA 540  
 CCGATTAAAA TATCAGGTCG TTCTTTTTTC AAATrCTTCA AATTACCTAT TTCACTAAAC 600  
 ACTTTGTTCT CGGCATTTTC TCTAATCGCA CATGTATTAA TTAAAATAAC ATCTGCAGTG 660  
 10 TTAATATCAG TCGTkGCTTG aTAGCCTAAT GCyTcmAGTA TACCAGCAAT GACCTCAGTG 720  
 TCATGTGCAT TCATTGACA TCCATATGTT TTAATTAAAA ATGTACGCTC GTTCCCCATA 780  
 15 CCGCGATATT TTyCATCAAT TtGGgAAATC nCTATTATAA CGAACTTCTk GTtTACnCC 840  
 TTTTTThCGC TCCTTTAAAA TTAAGGCGGC TGATAAACAG GTCCAAAATA TTAC 894

## (2) INFORMATION FOR SEQ ID NO: 570:

- 20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 441 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 570:

30 TATCAATCCC ACAGCACATG CTGAACAAGA TCAAACATGG GAGAAGATTA AAGAACGCGG 60  
 TGAACCTAGA GTGGGTCTTT CTGCAGATTA TGCACCAATG GAATTTGAGC ATACAGTTAA 120  
 TGGTAAGACT GAGTATGCAG GTGTAGATAT TGATTTAGCT AAAAAAATTG CGAAAGATAA 180  
 35 TAATTTAAAA TTAAAAATCG TCAATATGTC ATTTGATAGT TTGTTAGGAG CTCTTAAAC 240  
 TGGAAAAATT GATATTATTA TTTCCGGAAT GACTTCAACG CCTGAACGTA AGAAGCAAGT 300  
 TGATTTTCA GATTCATATA TGATGACTAA AAATATCATG CTTGTAAAGA AAGATAAAGT 360  
 40 TAATGAATAT AAAGATATcm AAGACTTTAA TAATAAAanA GTnGGGGCAC AAAGGGACTG 420  
 AACCAGAAAA AATCGCTCAA C 441

## (2) INFORMATION FOR SEQ ID NO: 571:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1205 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 50 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 571:

55

	AAGTCGTTCA TCTTTTTCTA CAAGGGTGTA AAAATAATCA ATCATATCGT ATAACGCTTC	120
	TTTACCAATG ATGTCATATG GTGTTGTTGT CATTTAATCA CCCATTTTCA AAAATTTACT	180
5	GTTACGAACT TAAGTTAATA TATAACTAAT ATAACATGAT TTAAACATT TGAAAGAAAT	240
	ATGCATATTT GCCAATTTAA TTTATATTGT TTGAAAGTGT TTCTTTTTTC TTGAAAAAAC	300
10	GTTGAACTTT ATTTAAAGGT tGATGATGTT CGAGGTTTAG TTCGTTtAAT AAAGATtGGA	360
	ACTTTTGTAAC CCTTGATTA TAGTCTTTAA CTTCGAACTC TAACTCATAA TCCGTAGTAT	420
	CGAAATACTC ACTTTTATCT AAAACCAGTA AATCACCTTT ATATTTAGTT TCTTGGCGAT	480
15	ATGTCGTTAA TGCACCAAGT ATTGATAAAG TTGTATCTTT TACACCAAAC TGTTCAACTA	540
	TAATTTGACG AATGTCATCT GGAAGATTGT CGTTTGAAAT AATCAAGTTC ATCTCTGGTT	600
	TAATGTCGAC GATATAGTTG TATTCTAATA GACCAACCTT TGCTGGTGTC TTTAAAGTCA	660
20	TTTCATATTG ATTGTCTTTA ACTCTTATGC GTAGTGCAGA GCGATGTTCC TTTAATTGTA	720
	AATCGGGTGT ATCAATATAG TAATTGACTT GCTTAmAAAG CACACTGTCT TTAATAATT	780
	TCTCTTGCAA TTTATTATAG ATTGAtGCAG TTATCATTTG TtTAAATTCT ATtTCATGAT	840
25	TTGTTGCCAT GATATGTATA CACCTCGTAT CAAATTCAAT TTATCTTAAC TATATTATGA	900
	ATGACAAAGT TGAATTTTAA AAGTAATTTT CTTTATCTAT TATCAATGTT AATTTGACCA	960
30	TTAAAAATAG TGTTGTAAG TGTTTGTAT TATTGaATTG TGTTAAATG TTATGGAATA	1020
	AGAGGAGGAT TAAGCATGsG TTTWTATATT AATGAAATTA AAATTAAAGA TGACATACTT	1080
	TATTGTTATA CAGAAGATTC TATTAAAGGA TTATCTGAAG TAGGACAAAT GCTCGTTGAT	1140
35	AGTGATAATT ATGCCTTTGC GTATACATTA GATGATGGTA AAGCGTATGC TTATCTCATT	1200
	TTCGT	1205

## (2) INFORMATION FOR SEQ ID NO: 572:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 570 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 572:

50	TGAAGAAATA GCAATGATGA AATGCCGTAT GTTATATGAG ACGGGGTCAT TTCTTGAATT	60
	AAGAGAAGAA ACAATTGTCT TATTGAAAAC TGGCATACAA CAATATGATG CATTGATGAT	120
55	TTATTACGTG AAAAGTTTGA TTGGTTTGGG ACAATATTTT GAAGCGGTAG AAGTAATTCA	180

ATTTGCTAAG TCAAAATTAA TTGAAGATGA AAAACGATTG ACTCAGTCAT TAGCTGATTT 300  
 TGtTACGTTA TCAATGAGGG AACAGACGCA CTTGATTTTG AAGTTAATAG ACAATGGTCA 360  
 5 TTTTCAATTT CAAGAAACGG TATTATATAT ATkAAaAyCT AATaCGTACa GTtATAACCT 420  
 CATTAGTTTA ATGATTGAGT ATTTAAGGTT CGCAAATTGT ACACAAGAAC TGACAATTGA 480  
 AAAGTATGGT ATGGATGTAA CTTTTGTACC AGCTAATTTA AAAGGGCTAG AACATACAAC 540  
 10 ACTTAAAGAA AAAGTTATAC CTAACGTTAT 570

## (2) INFORMATION FOR SEQ ID NO: 573:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 939 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 573:

GTTGAATGGT TAGCAGCTGC AGTTGTATTA TATTTCTGTG GTGTAATTGT TGACGCTCAT 60  
 25 GTATCATTCA TGTCTTTTAT TGCAATATTT ATCATTGCTG CATTATCAGG TTTAGTCAGC 120  
 TTTATTCTCG GTGGTTTCGG CGCTTTCGAT TTAGTTGTAT TACTAGGATT TAAAACCTTA 180  
 30 GGTGTCCCTG AGGAAAAAGT ATTATTAATG CTACTTCTAT ATCGTTTTGC GTACTATTTT 240  
 GTACCGGTAA TTATTGCATT AATTTTATCA TCATTTGAAT TTGGTACATC AGCTAAGAAG 300  
 TACATTGAGG GATCTAAATA CTTTATTCCT GCTAAAGATG TTACGTCATT TTAAATGTCT 360  
 35 TATCAAAAGG ATATTATTGC TAAATTTCCA TCATTATCAT TAGCAATTTT AGTATTCTTT 420  
 ACAAGTATGA TCTTTTTTGT AAATAACTTA ACGATTGTkT ACGATGCTTT tATATGATGG 480  
 AAATCACTTA ACGTATTATA TTCTATtGGC AATTCATACT AGTGCTTGTT TATTACTTTT 540  
 40 ACTGAATGTA GTTGGTATTT ATAAGCAAAG TAGACGTGCC ATTATCTTTG CTATGATTC 600  
 AATTTTATTA ATCACAGTGG CGACATTCTT CACTTACGCT TCATATATTT TAATAACATG 660  
 GTTAGCTATT ATTTTGTTC TGCTTATTGT AGCTTTCCGT AGAGCGAATA GGTTGAAACG 720  
 45 CCCAGTAAGA ATGAGAAATA TAGTTGCAAT GCTTTTATTC AGTTTATTTA TTTTATATGT 780  
 TAACCATATA TTTaTTGCTG GAACGTTATA TGCATTAGAT ATTTATACGA TTGAAATGCA 840  
 50 TACATCTGTA TTGCGCTATT ACTTCTGGCT TACGATTTTA ATCATCGCTA TCATCATAGG 900  
 TATGATTGCA TGGTTGTTTG ATTATCAATT TAGCAAAGT 939

## (2) INFORMATION FOR SEQ ID NO: 574:

55

(A) LENGTH: 1059 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 574:

10	GAATTAATTA AATATTACAC ACAGCCTCAT TTTTCATTTT CAAATAAATG GCTATATCAA	60
	TATGATAATG GAAACATTTA TGTGAACTT AnGAGATATT CATGGTCAGC ACATATATCT	120
	TTATGGGGCG CTGAAaGTYG GGGAAATATT AATCAGTTAA AAGATCGTTA CGTAGATGTG	180
15	TTTGGAATAA AAGACAAAGA TACTGATCAG TTATGGTGGT CTTATAGAGA GACATTTACA	240
	GGTGGCGTTA CACCAGCCGC AAAACCTTCT GATAAACTT ATAATCTTTT TGTGCAATAC	300
	AAAGATAAAC TACAAACGAT TATTGGTGCG CATAAAATAT ACCAAGGCAA TAAACCAGTA	360
20	TTAACATTGA AAGAAATCGA TTTCCGTGCA CGAGAAGCGT TAATAAAAAA TAAAATATTA	420
	TATAACGAAA ATCGTAATAA AGGTAAGCTT AAGATCACCG GTGGCGGTAA TAACTACACT	480
25	ATTGATTTAA GCAAAAGATT ACATTCAGAT CTAGCAAATG TTTATGTTAA AAATCCTAAT	540
	AAAAATACTG TTGACGTCCT CTTTGATTAG TATATGAAGG TGACTTATAC TTCATGCACT	600
	TTAATTCCAA ATCAGATTAT TTAAATGATA ATTTTAAAG TGTATGATGT ATATAATAGG	660
30	TAAAATTTTC TATATATTTA AATGGAATTG GGAGTAGGAA TGTGACAGAA ATAGTATTTT	720
	ATAAAATTTA TTCtTGTCAC TCCCCAACTT GCACATTATT GTAAGCTGAC TTTCCGCCAG	780
	CTTCTATGTT GGGGCCCCGC CAACTTGCACT TGTCTGTAGA aTTTCTTTTT GAAATTCTCT	840
35	ATGTTGGGGC CCCGCCTATA ATTGAAAAAT GCTTGTTACA TGGGCATTTT CATTGGGTCA	900
	ACTACTACCA ATATAATATT GtAGaGCCTA AGACATTGAT TTATTATGTC TTAGGCTCTA	960
40	TTCTTCATT TAATGATTAA nTTATTATAG CAATACTTTA TTGTCCCATG ATTAGTGTTT	1020
	TTTTAATGAG ACATAGTAAC TATAAAGTTT AATAATCGT	1059

(2) INFORMATION FOR SEQ ID NO: 575:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 574 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 575:

55	GTTTGCTTTA GGTCTGTTT CATATTTATA CTTCGAAGGA TTTACCTTTT TGaAGTCTGG	60
----	--	----

TTCAACTkGC TTTTATTCT TTTGAAATC AGCTGGTTGA GTAGTTATGA GTTCATTATT 180  
 TTTATTAGmA TAAATCTTAC CATTAAACATA TTTATAATCT TTTGTTATAA AGTCACCATT 240  
 5 TCTGAATGGA ACTACTTGAT TATGACCTTT AGAGAATAAA TCAGTACCGA ACATTAAATA 300  
 GTTCTTCGTA TCTATACCAG CCAAATGTAA AATTGTTGGC ATTACATCGA CTTGACCAGC 360  
 ATATTCAATTA TTGATACCAC CAGATTTACC AGGGATTTTA ATCCAGAAAC CAGTTCTGTT 420  
 10 TAAATCTGTA AATTTAGCCG GTGTGATTTT TTCACCTAAT AGTTTTTCCA TGGCATTGTT 480  
 ATGGTTTTCA GAGATACCAT AGTGGTCACC ATAAATCATA ATCACTGAAT TGTCATATAA 540  
 15 TCCTTTTTTC TTCAAGTCAT TAATATATTC TTCT 574

(2) INFORMATION FOR SEQ ID NO: 576:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 796 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 576:

CAATGTTTTA TAGTACAATA TATTTTnAAT AATACTCGTT AAGGAGAATG ATATGATATC 60  
 AATTCACGCA ATTTGACAG GAAAAATCCA AGATTTGCCT TATAGCTCGA AAAGACCGAT 120  
 30 GCGCTCTGCT TTAGATAAAA CTAAGATTTC ACAAACAACA TGGTTATCTT CAACTGGTTT 180  
 CACTGGTGAT GAACAGGCTT ATAAAGATCA TGGTGGACCA CATAAAGCAG TTTGTGGGTT 240  
 35 TAGTAAGCAT AATTATGCAC TGTATCAAGA TGATTTACCT ACACTACCTA CTCATGCGAT 300  
 GTTTGGAGAG AATTTAACAT TTGATTATTT AGACGAATCT GATGTTTACT TTGTAATCA 360  
 ATATCGTTTA GGTGAAGCGT TAATTGAGGT TTCTGAAATT AGAGAACCAT ACTGGAAAAT 420  
 40 TCAAGCAAAA TATAATATTC CTGATTTAGT GAAGCGCATG TCTACATCTG GTAAAAACAGG 480  
 TTTCTATTTT CGGGTATTAA AACAAGGCTA TGTATCTCCA AATGATCAGC TTTACTTAAT 540  
 ACAAGAAGCA CCAATCGAAC ATCGTTTATC TGTACAACAG CTTAATGACC TTTATTATAA 600  
 45 TGATAGACAA AATCAAGaTA TGTTACGATA TGCACTAAAC AATCCATTTT TGTCACCAAC 660  
 AAGACGCGAT AACTTCAAA AAATGTATAA CAGAACATTG GAAATAATTA CCTTTCATTn 720  
 50 ATAAGTGTTA AATGAACTTT TCAAAACAnA AAGGAATCAA CTTCACACAT CGTTTGTATG 780  
 AATAGTCTTA TCTATA 796

(2) INFORMATION FOR SEQ ID NO: 577:

55

(A) LENGTH: 1095 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 577:

```

10 AGAAATTATG ACAAATATA AAGATGGAAA GTTAGTTTAT GCATCAGTCG AACCAGGATC      60
   TTACGTAATA CATAAGATG ATGCAATTAA ATATGACGAT TATTCTAAGT TAAAAAATT      120
   AAGTCAGCTA ACTAACTTG ATCATCCAAA ACCAGTTCCA TATAGCGTaC TCAAATCAAA      180
15 TCTTTCGGAG TACCTTTAAC AAGCGTTTCA TTTATGACAC ATGGATCAAA GGATACTAAA      240
   GATGAAGTGT TGCCGGCATT GGCCTATTTT ACTTTTTTCAC CAAAAAATTA TGAAGACAAG      300
   TCTAATCCAG ATCCAAAAGT TTAAATTTA GTACATATGG ATTTCTTAAA TGCATCTAGT      360
20 GATTTTGGTA ACGCACATTT TGTTGTTTTA AGTAAATATA TTAAAGAGTA TGAATCAAAC      420
   TATGAAACAG CGTCAGATGA TTCTTTAAAA TAGTATTTAC TGTGTGAAAA ATAAATAGTG      480
   TACTACATTA AATAATCGCA ATAATAATCC CGATAAACAA TCAGCATTAC TGCTTATCAC      540
25 ATAGAGTTCG TAATAACTAT AACTCTATGA TTCGCAAATA ATAAATGATT GTCATCGGGA      600
   TTTATTTTTT TCAATTTATA AAGTGACATT ACCTTGTTCA TCAGCAGGTT TGAAAACAGT      660
   AATCACTGCA CTAATAATTG CTAAATGTG TGGGATACCT GTCCAACAGA ATATTAAGTG      720
   TAGAATACCT TGCATATTCT TGCCGGCATA AAATTTATGA ATACCAAAC TACCTAAGAA      780
   CAATGCTAAT AAAATATAAA TAACTTTGTT TACTTGCAAT TCTTCCCTC CAGTTGAATT      840
35 GCTTATAATG ACATTAGCTT CTCTTTTTAT TATACCCACT TTTAGTTCAA ACATTCTAGT      900
   TTAAGCATT CCAATCATCT AAATTTAGT TATTCAATCC TTACAATAAA TTTAGGATTA      960
   CATTTAGTTC GCATTGTATT ATTTACGTG TGAAATATAC GTAATGAATC ACATGACAAY      1020
40 CTyCAAATTG AAAAATATAC ATTCTATGAT GTAAGGTCGC ATTTTAATA TATTACGTn      1080
   AAAATAGTTT GGATG      1095

```

45

## (2) INFORMATION FOR SEQ ID NO: 578:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 489 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 578:

55

ACTGCCAAAA TAATCATAGC CATGTTCTAC AGCTGCTTTC GCTACAATAT CCAAACGCAT 120  
 TTCAAAACAA GCGGTACAAC GTAAGCCGCC TTCTTTTCA TCAGCTAATT CTTTATCCTT 180  
 5 CACCATTTTC ACAAACCTTAT GTGGTTCATA AGGTGCTTCA ATATACTTCA CATTGCGACC 240  
 AGTCTTGCGA TTAATACTTT CCACAAATTG TTCTTGCACT TTAGCACGTC GTAAGTACTC 300  
 ATTTTTCGGA TGAATATTTG AATTCGCGAA ATAAATTGCA ATGTCTGCAT ATTGTGTTAA 360  
 10 AAACTCTAAT GTATATGTAC TACAAGGTGC ACAACAATA TGCAATAAGa TTTTAGGTCT 420  
 GATTGCTTCT CTTTCCCACT GsCCGATTAA TnTCTTCAAC ACCTTGTTCA TAATTAATTT 480  
 15 GTnGATTn 489

(2) INFORMATION FOR SEQ ID NO: 579:

(i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 1287 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 579:

CGTACTGCAT CATTTGTTGA AAAATTTGCA AACTTCTTCA AGATGTTATG GCTTAGACTT 60  
 30 AAAGCGATGA AGCACTACAA AGCCTTAAAT AAAGAATCTA AGAAGCAAGA ATTTGAAAAT 120  
 TCATTCAAAG ATGTTCAAAA AATTATGCGT ATTGTGAATC ACAATATTAT TTTACGCTTA 180  
 AAAGAAGAAC AAAATAGTAC AAATGTACTT GAGGTTAGCT TAGTCATTAA TCATTACTAT 240  
 35 GATATGAGTC GCTCATTAAA GTGGCGTGCA CAACGTCGAA AAGAACGTCA AGAAAACAGC 300  
 AATCAAATCA TACCGCAAGC TATGTTCCAT AACCACAAAT TGGAAGCATT GTACTTACAA 360  
 CGTCATCTTT TAGATGAATT AATTCGCAAA AATAAAATCA ACAATATCGT TGCAGCTCAA 420  
 40 ATTCGAGAAA ATATCAATTa CAACGAAATT GTCTTGCTT TACAGTCCAA ACATTaAGCA 480  
 AGaCwTaCaw TmCCCCGTA CATATGAGAC AAAGTCATTA TCATCTCATA TGTACGGGGT 540  
 TTTTATATTC AACATCAAAA AATCAGATTG ATGAAAAGTA AATAACCTTT CATCAATCCG 600  
 45 ATTTGATTAT AGAATCTATT TTTAAGTTT AAATGGAATT GTACATACGT TAACATTCTT 660  
 TTGATAAATT AAATACAATT TCATACGCAA ACTAGTTTGA TTGTGTAATA AATTATGCCA 720  
 50 ACGTTTCTTA GTAATAAATT CTGGTATCAC TACTGTAATC ATATAGTTTT GATCGTTGGC 780  
 TTTACGATTA ATCTTATCGA TAAAACGTGA AATTGGTCGT ATAATACTGC GATATTCCGA 840  
 ATGTAAATA ACTAATCTTA CATCTGGGAA ATGACGTTTC CATTTCTCTT GGAATGCTTT 900

55



TCGGTAATAA ATAGACTTAT CAACTGCTGT TGTAATACTT GTTATCGGCA CAATTGCTAA 1020  
 ATTACGATCT ACCACGTCCA CATTAGAAGC ATCAATGTCA GAACGTAATT GTTCTGCGAT 1080  
 5 ATCTCGATAA TGKTTGkTAA TTTTCAAGAA GAAAATCACC ACGAACGGCA AGAAAATAAG 1140  
 TATCGGCCAT ACTTGGCTAA ATTTAGTTAT GAGTAAATC cATAAnAACA ATAAATGTCA 1200  
 CGATACCACC AAGTAAGTTC ACAGACAACT TACTTAACCA ATTCTTAGGA CGTTCATGAA 1260  
 10 TCCATTTAAT AACCATACCG AATTGTG 1287

(2) INFORMATION FOR SEQ ID NO: 580:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1223 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 580:

GCCTCCTTTC CTATATCCAT GGGTGCCTA GTTGGGAATG GCTTATTCCC TATCCCAAAC 60  
 25 ATCTGTCAAT TCAAGACTAT CACTGTATAC TAAATCGCCT nTCTATACnT CCATnCATTT 120  
 CATTTTCCAG TTGGAACAAT TTGTTTAATT ATTTTAAATA TCATTATCAA TCCAGAAGTA 180  
 TTTACTATTC ACTTTTACAA TAATCAATCA TTTAACTACA CATGGGTAGT CGGTGGTTTG 240  
 30 CTTGGGGTTA GCTTTTAAAC TGGCAATTTA TTGTTATTGC CAAAATTAGG TGCAACATTA 300  
 ACTGTAATTG CAACAGTTGC GGGTCAAATT ATTATGGGTG TCATTATTGA TACATTTGGA 360  
 35 TTATTTGGCG CTACAATTCA TGATTTTAAT TTAATTAAAG CAATTGGAGT ATTGTTACTC 420  
 ATTGTCCGCA TCGTCATAAT GAATCAATTT AACAAGAATA ATTTATTACT AACTGATCAA 480  
 AAGTATTTAC TGTTTTGGCT TCTATTAGGA TTTATTTTGT GTTCTTTTCC ACCTATTCAA 540  
 40 ACGACAATTA ATAGTGCTTT AGCTAGTCAT ACTCATTAC CAGCCTTTGC ATCATTAGTA 600  
 TCATTTACAA TTGGGTCAAT AGCGCTATTG ATTTTAAACG CTATTTTAA TCGTTCTTTA 660  
 AAACTAAAAA CAAGTCATTT AAAATTCGGT AAATTAAAGC CTATCTATTT TACTGGCGGT 720  
 45 ATACTTGGCA TGGCTTTTGT AACAGCTAAC ATTATCTTAA TGCCTCATAT GGGTGCAGCA 780  
 TTAACAACAC TTATTGGGAT GTTTGGCCAG ATTCTAATGG GCATATTGAT AGATCACTTT 840  
 GGATTATTTG GTTCACCTAA AATAGCAATG ACATCCAGAA AACTATTGG TCTATTATGT 900  
 50 ATTTTGACAG GCATTATACT TTTAAGATTA TTTTAAATTA ACTTTTAGCT TATCATTTTA 960  
 ACTTGTAATT ATTTTAAAAA GTGATAAGCT ATTTTTTTGT GGTCTAAAAA TCTTTAGAAA 1020

55

CAACTCATT C TTAAGACCTA AATTAATGTT ATnTTTTAAT AATTTACACC AAATTAATAG 1140  
 CAAAAATTAT GTTATTCGTG CTAATATTTT ATAGTTGGTT ATTCAATTAA TTAAAAATAA 1200  
 5 GTCAAAATGC ACAACTTTTT ATn 1223

(2) INFORMATION FOR SEQ ID NO: 581:

(i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 454 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 581:

CACCTTTTTA TCATGCTTAG TTATCAATAA ATCTATATTT TGCTGTTTTA CAATTTTTTT 60  
 20 AACTTTATCA ATCTCATTAT CTTGGACTAA ATAAATATAT GATCTTGCAT CTGTTGCTAG 120  
 AGCTTGTTTCG TGTTTTTCTG ATAAACATA TGTGATGGAA GCGTGAATAA TAATGCCTAA 180  
 TGTAACAAAA CTGATAATTA ATATACTGCT TATCAATAAC ATTAAGCGGT GGTGAAACTT 240  
 25 CATCATTGTT CTTTAGGTCT TTCCAATTTA TAGCCTAAGC CACGCACAGT TTTAATAAGT 300  
 TGTGGCTTCT TAGGATTATC TTCTAATTTA TCTCTTAAAT GACTGATATG TACATCAACA 360  
 ATTCTTGAGT CTCCTGCAAA TTCATAATTC CATACCGTAT TTAACATATG CTCTCTCGTA 420  
 30 ATGACTCTGC CTTGTCTTTC TATCAAATAA AGCA 454

(2) INFORMATION FOR SEQ ID NO: 582:

(i) SEQUENCE CHARACTERISTICS:  
 35 (A) LENGTH: 452 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 582:

TTTATAAGAT TTTATTTACA AATAATTGGT TTTCATATGT ATAAACACTT TTGACTTTCA 60  
 45 AATCTTAACG ATAATTCTAT TACAATACAA TCCCCTATTA GAATGATTTA TGTAATAAAA 120  
 AAAGCGGAGT TTCCCCCAGC TTTTCTAAAC GACTACATAA AATATAAGAT TGCAATTAAA 180  
 TGCAATAGTG ATGCTATTAC AATAAAAATA TGCCAAATCA TATGAAAATA TGGTCTATTC 240  
 50 TTTTGTGCAT AAAACCATGC ACCAATTGTA TAAGACACAC CACCTAAGAA AATGAATAAT 300  
 ATGAATATCC ATGATGTGCG AATAAAAATA ATTGGTAACA AGATAATACC TACCCAGCCC 360

55

AAAATCCCCC AAAGTGTGCT TCCCCATAAT AA

452

(2) INFORMATION FOR SEQ ID NO: 583:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1472 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 583:

15	CGCTTTTTCG AATAAGTCAT TAGCCGCTTT TAAGCCTTCT TCTTTTCGAT CTACAACAAG	60
	TAAAATAAAT GGCTTTAACG CTTCTTCTTT TTCACCTTCA AGCATATCTG GTTTTTGAAC	120
	CATTTCAAAT GGAGATTTCA ATCCATTATT ATCGCTCATT TCAATAATTG CATCATACTG	180
20	TGCTTGTGAC ATACTTGCAA TAGCCTGTTT TGCATTTTCT TGAAGGAAAT ATAAGTTTTT	240
	CAATTTAGGA TGCTTATTTA ATGTACTTAA TGTAATCGGT GTAATGTCTT TCTCATAAGA	300
	CACTTCAATC ACTGTACTAT TTGTTCTACC AGGAATTGGT GGTTTTTTCAT GAATATGCTT	360
25	TGATACTTCT CCAATTCCAA CGACAGATTG ATTTTTCGTT CGATTATAAA AAATAATATT	420
	GTCGCCTTCT TCTAACTGAG TATAAAAATG ATAACCATTA CGTTTAATAC CGTTGTACGT	480
	GTGCGTATAA ATCGTATATT GGTTCACAGG TTCAAATTCT TCAGTTTCAG CTAAAAAGAA	540
30	ATAACGCGGT ATCTTAATTT CGCCTTTACC AAGACCACTT ATTAATCAAA ACTCTTCTGC	600
	AGTGATTTGA TTGAACAATG TCTCTTTCAT ATTAATTATA CGAAATTCCA AAGCTTCACT	660
35	ACGCTTTAAA TAATCTGCTG TTAATGGTTT CAATTGTTCA TTAACACGAA ACTGTACACG	720
	TATTTTATTT TGTGCACCTG TTTCAACACT AATAATTTCA CCACATCCAA GTAGTCCAGT	780
	ATCCGTCTGA ACTTGATAAA AGATGACTTG ATCTCCTACT TTAGCCTTTT TAAACGCTCT	840
40	AAATCCTTGA GATGGGTAA AATGTGCGCC TGATTCAAAT AAAGCTGTTT GTCCTACTAA	900
	CGGTCATTA TGATTCCAAC GGTATATACC ACAATTCAAC CAAAATAAT TCGTTTCTGC	960
	TGTCATCTTA ATACTCCTTA ACCTGAATAA ATTTTAGAAA CACTATGAAT TACATTCTTT	1020
45	TAGTGTCTTCT TATGCAGTTG GACGCGTATG CGAACAACTG TATACCCTTT GTTCACTGCG	1080
	ATTTAATCG CATTTCTAT AACATTGTAG CGCCAGGAC ATTAATTTAC GTCCAGACC	1140
	CTTATCGTTT TCACTTCTAA GTAAGTCGAA CTATTTTGCT TTACAACAAG TGCGACTCTA	1200
50	AATACAGTTG GACACACATA CGAGCAACTG TATACCTTTT AATCAGTTTT CTATATTTTA	1260
	TTTATTATAT CTGTCTTAAT GATAAAAATT GTTACAAACA GTTTAACATA TTTAGCTACC	1320

55

ATGATACCAC TATGCTTGCh TATCTCTATA GCGCCATTGA TACACATTTT TAAATATCTA 1440

TACTGCCGTT AGAATTTTAT CATGTCThAA TT 1472

5 (2) INFORMATION FOR SEQ ID NO: 584:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 787 base pairs  
(B) TYPE: nucleic acid  
10 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 584:

GGTAGTGAAT GGGGTTCAAG ACAACAGTAT TGGAGTACG AATGAATCAC AGTTTTTAGG	60
AAATTATATT GTAATTAAGC ACGCAGAAAA TGAGTATAGC TTAATAGCTC ATTTACACCA	120
20 ATATTCAATC ATTGTGAATG AGGGGCAAAA TGTTAAATAT GGTGATATCA TTGGGAAGGT	180
TGGGAATTCT GGCAATTCTA CTGAACCTCA TATACATTTT CAAGTAATGA ATGATAAGAA	240
TATTGAAGCA TGTACATCTT TAAAAATTCTG ATTTATAAAT AATCGAGAAC TTATCAAAGG	300
25 GGATGTGGTC TCGCGATTAC AAGCTGAATG ATGGCGATAC TTATAAAATC TCGACACTAT	360
AAAAATGGTA TAGTGTGAG ATTTTCTTGC TTATTAGTT AATTCAAAGT GCACGCCGGA	420
TTCATTAGAA GTCGACGTAT TTTTGTTTGT AATAGAGTAA CCGGTCATTG AAATTTTAGA	480
30 TTCAATATCT GAAGCGGAAT TTGTAGATTC AGGATTATAG AAGCTACATT CATAAGTGTT	540
ATCATCTTTC TTTTAAAGTA TAAACATACC TTTGGCTTTA ACTTCGACTT TAGTGTGTGT	600
AATGTCAAAA GTTTGAGTGC TATTACTATA ATTAACACCA GCCCAAACCG ATTCATTATC	660
35 TTTCACAACG GGAAGTCAT CTCTTGCTT AACGACGThA CTTCATCTTT CTCTGTCTTA	720
AAGACATCTT TAGATAAGCC TGGATACAAC ACATATCCAT ATTTATTGTC AGAATTAGAA	780
40 TGCTTTT	787

(2) INFORMATION FOR SEQ ID NO: 585:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 830 base pairs  
45 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 585:

ATGCTAAATT AGGGGAATA TTTACGATAA AGAGACCAGA AAAATAATAA AACACATGCT 60

55

TAAAATGAAG ACTATTTTTT ATTACAAGAA AATGTATCTA GTAAACTTAA AGTAGCAAGA 180  
 CCTAATAAAT TTAATGCATG TTGTGCACCT TTTTACCTT GGCCAGCTTC GAAATGTTTG 240  
 5 TAAGCAGCTA CACTTAAAAT GCCTATCGTT GATAGTGATG CAAGGCGAGA AATGTTTTTA 300  
 TTGATAAAGC TAGCTGAGTA TAAAGCAGCA GTAGTTGCTT CTGCAATGCC GACGTATTTT 360  
 ACAAGTTCTT TTTGCAAGCC AAAAGTATGT TCAAACAGTT CAATCATACC CTTATCTTCT 420  
 10 TGCAATTTAG GTTTACTGGC TTGGTATAGC TCTTTCGCAA GTTTTAAATT CGTTGCGTAA 480  
 CGCAAAATCA TATTTAATTC CTCCAATAT TTGATTTTTT GTGAAAGATG ATTACTTTAT 540  
 15 CATTTTTACC CGTTTCTATA AAAATGAATC AATTATGTAA CGTATGTGTA GTTTAGGAAT 600  
 GTTTGCTATG GAAATATAAT TCTGTTCACT CAAAATGTAT GAAATTAATG TGTAGTTTGT 660  
 TCGAGTTGCT CTTTAAATTT GGTTAGATTG TTTTITAGAG AAGCGGTACT ATTTTAAAGT 720  
 20 GCATCAACAG ATTTACCTTC GTTTGAGAC ATTGAGTTTA TTACAGCACG AAGTTCGTGT 780  
 TCTAGTATGT CaGCGTCGct TTAGCATTAG AACTTAaTat TTAtAcTCTT 830

## (2) INFORMATION FOR SEQ ID NO: 586:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 412 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 586:

35 TTAGGACGTT TTTACAATCA GTACAATGAG CTCATTGTTA TTAGTCCTTT AACGGCGTCT 60  
 TTTAATGCTG GCGCTACATT TGGGCGATT CATCATTTAA TTGATACTGA AACTTTAGCA 120  
 AAATTAGAAC ATGAAAAAGG ACATTATTAT CAGAAGATGA TATGTGATGA CAATGTAGAA 180  
 40 ATGATTTCTA TAAATAACAT ACCGAAATAT CCGAGAAATC ATAATGTATT AACTAATCAT 240  
 GACTCATACG AATATTCATT GAATTTAGGA AGTAGTAATA GTTATTCAAA GTATGAGCTT 300  
 ACCTTAGATG ATATTTATGT TGGTGCTACC TTTtAACAAA TTATATTTAT ATTCTAGCCm 360  
 45 ActAAATAAA AGGGkaCtaT TTGaATCaAA CmATaTgTAT TAACCTTTTT TA 412

## (2) INFORMATION FOR SEQ ID NO: 587:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4709 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 55 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 587:

	TTCAGTTTGA AATTAATCAT ATAAATTTCT TATGGGAGGG TTGATATCTT AATGATTAAAC	60
5	ATTATTTTCAG CTATAGGATC TATTGGAACA TTTATTATGG CTTTATTTTA TTTTGTATCA	120
	GTTTCAGTTC AACTTTATCA AATGAAAATT AGCTTTCTGC CAGCTTTAGG TTTTAACCAA	180
10	ATTTTATTAG AAAGGGAGGA GGATCAACTT AATATAATGA ATTCGGCAAC AGAAGAGCAT	240
	CATCATAAAG ATTATATTAA ACTATATAAT TTAGGTGGCG GTGCTGCTAA TAAAATTGCA	300
	ATAGAGGTTT TATTGGGGAA GGATATAGTC ATTCAGAAAA AATACGTGCA TATTkTACCT	360
15	AGTAAAGAAG GGTACATGTT ACCAATTAAT AAAATGTGT ACGAAGAATT AGAAAGAACG	420
	ATTGAsAACA ATGGTCATGA AGCTGATTTG AATGTACGTA TGACTTATTA TCATAATGTA	480
	AGTCGCAAAC AACAGGAAGT TATATTAAAA GGTCAAATCG ACCGTTTTAA TACTTATAAT	540
20	AATAAGAAA TTTATGATTT GCAGTTTATC TAAAAATTGA TTTAAGAGGG TAGTTGTTTA	600
	TTGCGAAAAA TATCATTCAA TTTTAATGAA ATAATGGCGT CATTACTATA AAATATTACT	660
	TTATGTTGTA ATGCATTTTT CTATAAGATA GAACTAAAAG GAGGGGCAAA GATGCAAATT	720
25	AGACAAATAC ATCAACATGA CTTTGCTCAA GTGGACCACT TAATTAGAAC GGCATTTGAA	780
	AATAGTGAAC ATGGTTATGG TAATGAATCA GAGCTAGTAG ACCAAATTCG TCTAAGTGAT	840
30	ACGTATGACA ATACCTTAGA ATTAGTAGCT GTTCTTCAAA ATGAAGTTGT AGGGCACGGT	900
	TTACTAAGTG AAGTTTATCT TGATAACGAG GCACAACGGG AAATTGGATT AGTGTTAGCA	960
	CCTGTATCTG TTGATATCA TCATCAAAAT AAAGGTATTG GGAAGCGATT GATTCAAGCA	1020
35	TTAGAACGAG AAGCAATATT AAAAGGATAT AATTTTATCA GTGTATTAGG ATGGCCGACG	1080
	TATTATGCCA ATCTAGGATA TCAACGCGCA AGTATGTACG ACATTTATCC ACCATATGAT	1140
	GGTATACCAG ACGAAGCGTT TTTAATTAAA GAATTAAAAG TGAACAGTTT AGCGGGAAAA	1200
40	ACAGGTACCA TAAATTACAC ATCTGCTTTT GAAAAAATAT GATTTCAGC TAGGATTACA	1260
	TTAGGTAGAG TTCATATTAA TAATAAAAAA TGTTTGCAAT CAAATCGTAC GTTGTCGTTT	1320
	GTAATTCCTA AAATAGCAAT AAATAAAATG TTTGTTAGTA AAGTATTATT GTGGATAATA	1380
45	AAATATCGAT ACAAATTAAT TGCTATAATG CAaTTTTAGT GTATAATTCC ATTGACAGAG	1440
	ATTAAATATA TCTTTAAAGG GTATATAGTT AATATAAAAT GACTTTTTAA AAAGAGGGAA	1500
50	TAAAATGAAT ATGAAGAAAA AAGAAAAACA CGCAATTCGG AAAAAATCGA TTGGCGTGCC	1560
	TTCAGTGCTT GTAGGTACGT TAATCGGTTT TGGACTACTC AGCAGTAAAG AAGCAGATGC	1620
	AAGTGAAAAT AGTGTTACGC AATCTGATAG CGCAAGTAAC GAAAGCAAAA GTAATGATTC	1680

55

	GTCAAACACT AATAATGGCG AAACGAGTGT GCGCGAAAAT CCAGCACAAAC AGGAAACGAC	1800
	ACAAATCATCA TCAACAAATG CAACTACGGA AGAAACGCCG GTAAGTGGTG AAGCTACTAC	1860
5	TACGACAACG AATCAAGCTA ATACACCGGC AACAACTCAA TCAAGCAATA CAAATGCGGA	1920
	GGAATTAGTG AATCAAACAA GTAATGAAAC GACTTCTAAT GATACTAATA CAGTATCATC	1980
10	TGTAAATTCA CCTCAAAATT CTACAAATGC GGAAATGTT TCAACAACGC AAGATACTTC	2040
	AACTGAAGCA ACACCTTCAA ACAATGAATC AGCTCCACAG AGTACAGATG CAAGTAATAA	2100
	AGATGTAGTT AATCAAGCGG TTAATACAAG TGCGCCTAGA ATGAGAGCAT TTAGTTTAGC	2160
15	GGCAGTAGCT GCAGATGCAC CGGTAGCTGG CACAGATATT ACGAATCAGT TGACGAATGT	2220
	GACAGTTGGT ATTGACTCTG GTACGACTGT GTATCCGCAC CAAGCAGGTT ATGTCAAAC	2280
	GAATTATGGT TTTTCAGTGC CTAATTCTGC TGTAAAGGT GACACATTCA AAATAACTGT	2340
20	ACCTAAAGAA TTAAACTTAA ATGGTGTAAC TTCAACTGCT AAAGTGCCAC CAATTATGGC	2400
	TGGAGATCAA GTATTGGCAA ATGGTGTAAT CGATAGTGAT GGTAATGTTA TTTATACATT	2460
	TACAGACTAT GTAAATACTA AAGATGATGT AAAAGCAACT TTGACCATGC CCGCTTATAT	2520
25	TGACCCTGAA AATGTTAAAA AGACAGGTAA TGTGACATTG GCTACTGGCA TAGGTAGTAC	2580
	AACAGCAAAC AAAACAGTAT TAGTAGATTA TGAAAAATAT GGTAAGTTTT ATAACCTATC	2640
30	TATTAAAGGT ACAATTGACC AAATCGATAA AACAAATAAT ACGTATCGTC AGACAATTTA	2700
	TGTCAATCCA AGTGGAGATA ACGTTATTGC GCCGGTTTTA ACAGGTAATT TAAAACCAAA	2760
	TACGGATAGT AATGCATTAA TAGATCAGCA AAATACAAGT ATTAAAGTAT ATAAAGTAGA	2820
35	TAATGCAGCT GATTTATCTG AAAGTTACTT TGTGAATCCA GAAAACCTTG AGGATGTCAC	2880
	TAATAGTGTG AATATTACAT TCCCAAATCC AAATCAATAT AAAGTAGAGT TTAATACGCC	2940
	TGATGATCAA ATTACAACAC CGTATATAGT AGTTGTTAAT GGTCATATTG ATCCGAATAG	3000
40	CAAAGGTGAT TTAGCTTTAC GTTCAACTTT ATATGGGTAT AACTCGAATA TAATTTGGCG	3060
	CTCTATGTCA TGGGACAACG AAGTAGCATT TAATAACGGA TCAGGTTCTG GTGACGGTAT	3120
	CGATAAACCA GTTGTTCTCG AACAACTGA TGAGCCTGGT GAAATTGAAC CAATTCCAGA	3180
45	GGATTCAGAT TCTGACCCAG GTTCAGATTC TGGCAGCGAT TCTAATTCAG ATAGCGGTTC	3240
	AGATTCGGGT AGTGATTCTA CATCAGATAG TGGTTCAGAT TCAGCGAGTG ATTCAGATTC	3300
50	AGCAAGTGAT TCAGACTCAG CGAGTGATTC AGATTCAGCA AGCGATTCCG ACTCAGCGAG	3360
	CGATTCCGAC TCAGACAATG ACTCGGATTC AGATAGCGAT TCTGACTCAG ACAGTGACTC	3420
	AGATTCCGAC AGTGACTCAG ATTCAGATAG CGATTCTGAC TCAGACAGTG ACTCGGATTC	3480
55		

CGATTCTGAC TCCGACAGTG ATTCCGACTC AGACAGCGAT TCAGATTCCG ACAGTGATTC 3600  
 CGACTCAGAT AGCGATTCCG ACTCAGATAG CGACTCAGAT TCAGACAGCG ATTCAGATTC 3660  
 5 AGACAGCGAT TCAGATTCAG ATAGCGATTG AGATTCCGAC AGTGAAGTCAG ATTCCGACAG 3720  
 TGAATCCGAT TCAGATAGCG ATTCAGATTC CGACAGTGAC TCAGATTCCG ACAGTGATTC 3780  
 AGACTCAGAC AGTGATTCGG ATTCAGCGAG TGATTCGGAT TCAGATAGTG ATTCCGACTC 3840  
 10 CGACAGTGAC TCGGATTCAG ATAGCGACTC AGACTCGGAT AGCGACTCGG ATTCAGATAG 3900  
 CGATTCCGAC TCAGATAGCG ATTCAGAATC AGACAGCGAT TCAGATTCAG ACAGCGACTC 3960  
 AGACAGTGAC TCAGATTCAG ATAGTGAAGC GGATTCAGCG AGTGATTCAG ACTCAGGTAG 4020  
 15 TGAATCCGAT TCATCAAGTG ATTCCGACTC AGAAAGTGAT TCAAATAGCG ATTCCGAGTC 4080  
 AGTTTCTAAC AATAATGTAG TTCCGCCTAA TTCACCTAAA AATGGTACTA ATGCTTCTAA 4140  
 20 TAAAAATGAG GCTAAAGATA GTAAAGAACC ATTACCAGAT ACAGGTTCTG AAGATGAAGC 4200  
 AAATACGTCA CTAATTTGGG GATTATTAGC ATCAATAGGT TCATTACTAC TTTTCAGAAG 4260  
 AAAAAAGAA AATAAGATA AGAAATAAGT AATAATGATA TTAAATTAAT CATATGATTC 4320  
 25 ATGAAGnAAc rCCTTAAAAG GTGGCTTTTT TACTTGGATT TTCAAATAT ATTGTTTGAA 4380  
 TATAATTAAT AATTAATTCA TCAACAGTTA ATTATTTTAA AAAGGTAGAT GTTATATAAT 4440  
 TTGGCTTGGC GAAAAAATAG GGTGTAAGGT AGGTTGTTAA TTAGGGAAAA TTAAGGAGAA 4500  
 30 AATACAGTTG AAAAATAAAT TGCTAGTTTT ATCATTGGGA GCATTATGTG TATCACAAT 4560  
 TTGGGAAAGT AATCGTGCGA GTGCAGTGGT TTCTGGGGAG AAGAATCCAT ATGTATCTAG 4620  
 TCGTTGAAAC TGAATAATAA TAAAAATAAA TCTAGAACAG TAGAAGAGTA TAAGAAAAGA 4680  
 35 TTGGATGATT TCAATATGGT CcTTCCCA 4709

## (2) INFORMATION FOR SEQ ID NO: 588:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1554 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 588:

CTTTTTTAAc TAwCgGaAtA TTGtCaTgaT tAcAcTTCGt TAGGGTTTAc gTCGtAATTT 60  
 50 CATTTaATAA gCgCTTCaCc ATTAAAtGTG gTAmCCTTTA ATTCGCCAGT AGAAACATCA 120  
 CAGTAACTAA GCGCAATTTC AGGTTGATTC ATAACAAAAC TTAAATATA GTTATTTTGT 180

55



CGTCTAACCA TACCTTTCGT TTGTTTCGGA TCTTCCATCT GTTCACAAAT AGCTACTTTA 300  
TATCCATTAT TAACAAGTGT ATCTATATAA CTATCTGCAG AATGATACGG AACACCACAC 360  
5 ATCGGAATTG GATTTTCTTT TTTAGCATCT CTTTtagTTA AAGTAATTC AAGTAyACGT 420  
GATGCCTCCT TGGCATCTTC ATAAAACATT TCATAGAAAT CACCTAGTCT AAAAAATAAT 480  
AAGCAATCTT GGTATTCTGA TTTTATTTTT AAATATTGCT GCATCATTGG TGTAACATTA 540  
10 GACATATTAT TTCTTCACAA CCCTTGTCTC TTTTAAAAAT TTGTCTTTAC AATATATTCTG 600  
TTTGTAAGyT TTTTAATTAT TAATTATTTA ACTTATACAT TTTAACATAC TTACTTTTAC 660  
AAACCTATTc ATACCATATA ATCACGAAGC ATCTTAAATG TATAAGAAAA CGCCTCAAAC 720  
15 CTAATAAAAT GTGTCAATAG CATGTTTAgA ATTAAATTAA AATTCTAACA TTCAAGACAT 780  
TTAATTAAGT AAGGGCGTTC AATATTAAAA TGAACAATGA CTCTGTTTGA AATCATATAT 840  
20 CATAAAATTA TTTTATAAAC CTTTGAAGAA TACCACGTTT TTTTAGAGTA ATTAATAAGA 900  
AATAACTTAT AATAGATCCG ATAGCACTTG aGACTATGAa CGTAATCATT AACGGTTTAA 960  
TGAAGAAGTC TTGAAGCCCA AGGaAATATG CTAATGGtAT aCaAATTAAA cTTCCgATGA 1020  
25 CaCCAGTTCC aAGTACTTCa CCGACCGCGG CCaTAAATAT ATGTTTACGA TATnygTAAA 1080  
ACATACTAGC CAATAAAACT CCAATCATAC TACCCGAAA TGCAAAAGst GTACCAGTAC 1140  
CAAAAAGAAC TCTTAAATT GATGATATAA GCGCTTGAGC TAATCCATAC CAAGGACCTA 1200  
30 CTATGACCGC ACTTAATACA TTTACAAAAT GCTGTACTGG TGCTGCCTTA ACTGGTCCTA 1260  
GAGGAATGAT GATAATACTG CTTAATACAA CATTtATTGC AATTAAAAGT GCAGTTATAG 1320  
CCAGTTTtCT TGATTTCATA TGATTGTTCT CCTTTTTGTT TGTAATTAAT CACTATGCTT 1380  
35 GGCTTTATTA TGGTCATTa AACGTGTTTC CATTGTTGAT ACAAACATTT TCAATAATTG 1440  
ATTCGCTTCA TATTGTGAAG TTTGAACTG TTCAACTATG GGCAATGTAT TTATTTCTGC 1500  
40 TTCTAtACTC TGAATGGtAT GTTCCGACTG nTCCAGCGCA TTTGTTTCCC GnAA 1554

(2) INFORMATION FOR SEQ ID NO: 589:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 638 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 589:

TTTTGACGAA AAAAGTTGAT TTACAAATAT ATAAACGTTG TGATTTCaAT GTTTGTATAG 60

TATATACATG ACAGCAACTT GGGAAAAAAA GGAAGGTAAC GAAGGTTTAT TAACTGTTAC 180  
 TGTTCCTGCA GAAAAAGTAA ACAAAGCTCT GAGAGATCCC CTCATAATTT CCCCAAAAGCG 240  
 5 TAACCATGTG TGAATAAATT TTGAGCTAGT AGGGTTGCAG CCACGAGTAA GTCTTCCCTT 300  
 GTTATTGTGT AGCCAGAATG CCGCAAAACT TCCATGCCTA AGCGAACTGT TGAGAGTACG 360  
 TTTCGATTTC TGA CTGTGTT AGCCTGGAAG TGCTTGTCCT AACCTTGTTT CTGAGCATGA 420  
 10 ACGsCCGCAA GCCAACATGT TAGTTGAAGC ATCAGGGCGA TTAGCAGCAT GmTATCAAAA 480  
 CGCTCTGAGC TGCTCGTTCG GCTATGGCGT AGGCcTAGTC CGTAGgCAGG ACTTTTCAAG 540  
 TCTCGGAAGG yTTCTTCAAT CTGCATTTCG TTCGAATAGA TATTAACAAG TTGTTTGGGT 600  
 15 GTTcGAATTk CAACArGTaA GTtAGtTGCT AGAnCCCA 638

(2) INFORMATION FOR SEQ ID NO: 590:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1242 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 590:

AAAATATTCC CGTACATTTT GATGTCTGTA GGGGCTTTTT TGACTTTAGG ATTTGTCATT 60  
 30 TTTTCAATTC ATAAAGGGAG ACGAACGAAA AATGAATCAG CACGTAAAAG TAACATTTGA 120  
 TTTTACTAAT TATAATTACG GCACATATGA CTTAGCAGTA CCAGCATATT TACCGATAAA 180  
 AAACCTAATA GCTTTAGTAT TGGATAGTTT GGACATTTCA ATATTTGATG TCAATACACA 240  
 35 AATTAAAGTG ATGACGAAAG GTCAATTACT TGTTGAAAAT GATCGACTCA TTGATTATCA 300  
 AATCGCTGAT GGAGATATTT TGAAGTTACT ATAGGAGGGA AAATAGATGG TTAAAAATCA 360  
 40 TAACCCTAAA AATGAAATGC aAGATATGTT AACGCCTTTA GATGCTGAAG AAGCAGCTAA 420  
 AACAAAATTA CGCTTAGATA TGAGAGAGAT TCCTAAGTCT TCAATTAAAC CAGAACATTT 480  
 TCATTTAATG TACTTATTAG AACAACTTC TCCATATTTT ATAGATGCTG AATTAACCTGA 540  
 45 ACTACGTGAC aGTTTCaAA TACATtATGA CATTAAATGac AATCATACAC CTTTTGATAA 600  
 TATTAAATCA TTTACTAAAA ATGAAAAATT ACGTTACTTA CTCAATATCA AAAATTTAGA 660  
 AGAAgTAAAT CGTACACGCT ACACATTTGT GTTGGCACCA GATGAATTAT TTTTCACAAG 720  
 50 AGATGGATTA CCCATTGCTA AAACAAGAGG GTTACAAAAT GTTGTGATC CATTACCTGT 780  
 GTCAGAAGCT GAATTTTTAA CAAGATATAA AGCGCTGGTT ATCTGTGCAT TCAATGAGAA 840

55

AACTAAAGTT ATTGAAGCGG CAACGTTAGA TTTACTAACG GCATTTTtag ATGAACAGTA 960  
 TCAGAAACAA GAACAAGATT ATAGTCAAAA TTATGCATAT GTACGCAAAG TAGGACATAC 1020  
 5 CGTTTTCAAA TGGGTTGCTA TCGGTATGAC AACGTTAAGT GTTTTATTAA TTGCATTCTT 1080  
 AGCCTTTTTTA TATTTTTCAG TAATGAAGCA TAATGAGCGC ATTGAAAAAG GATACCAAGC 1140  
 ATTTGTAAAG GATGtTATAC GCAAGTACTA AATACGTATG ATGATTTAGA TGGTaAAAAAt 1200  
 10 TgaTAAAGAG GCACTTTACA TTTATGCCAA AAGTTATATC CA 1242

## (2) INFORMATION FOR SEQ ID NO: 591:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 744 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 591:

TTCCAGATAG AGCCTTAGTT GCTGCCGCTG AATTGTCTGA TCGTTACATC ACTGATCGTT 60  
 25 TTTTACCAG ATAAAGCGAT TGATTTAGTT GACCAAGCAT GTGCAACAAT TCGTACGGAA 120  
 ATGGGATCAA ATCCAAGTGA ATTGGATCAA GTTAATAGAC GTGTCATGCA ATTAGAAATT 180  
 30 GAAGAAAGCG CACTTAAAAA TGAATCTGAC AATGCGAGCA AACAGAGATT ACAAGAACTA 240  
 CAAGAAGAGC TTGCCAATGA AAAAGAGAAA CAAGCAGCAC TTCAATCTCG TGTagAATCA 300  
 GAAAAAGAAA AAATAGCAAA TTTACAAGAA AAACGTGCGC AACTAGATGA AAGTAGACAA 360  
 35 GCGTTGGAAG ATGCACAAAC AAATAACAAT TTAGAAAAAG CTGCTGAACT ACAATATGGA 420  
 ACAATTCTCTC AATTGGAAAA AGAACTTAGA GAATTAGAGG ATAATTTCCA AGATGAGCAA 480  
 GGTGAAGATA CAGATCGAAT GATTCGTGAA GTTGTAACAG ACGAAGAAAT TGGCGATATT 540  
 40 GTCAGCCAAT GGACAGGCAT ACCAGTTTCA AAATTAGTTG AAACAGAACG TGAAAAATTA 600  
 CTTCACTTAA GTGACATCTT GCATAAACGT GTTGTAGGTC AAGATAAAGC GGTGACCTG 660  
 GTTTCAGATG CAGTAGTTAG AGCAAGAGCA GGTATTAAAG TnCAAACAGA CCTATTGGTA 720  
 45 GTTTCTnATT CCTAGGTCCn ACTG 744

## (2) INFORMATION FOR SEQ ID NO: 592:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1449 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 592:

	TTGTTATAGT GTCTGAATCT GTGTAACCTCT TGCCCATGTG TTCTGAGTAA AGCACCCACT	60
5	GTTTATTTAC TTTTCGTTGT AGTCTAGCTT CGTGTAGTAG TTGTTTAAAC TTTGTGCTG	120
	ATATACCGTA GTCTGCCGCG ATTTGAGTTG TGGCTAATGT GCCAGTTGAT TTAAAGATTT	180
	CATCAACATA ATCTGCTTTG GGTTTTAGCT CTCCAATTTT TGTTTGATAA AGTAAGTTTT	240
10	GCTCTTTTTC TTTCTTATAC TCAGTCAACA CTGTAATGAT GTAGTCTGGA TCTTTTAATG	300
	TTTGTTCAAT TACATTGTCT GTTGCGTATA TACCGTGTTC GCGAATAGCT GGTaGGACAT	360
	CTGATGTTAC CCAGCGTTTG AATTTTCTAG CGGTTTCTCT AATTTTTCG TTTTGCTTT	420
15	GTTTAGAAGC ATCGAAGATT AGACTGTATA ATCCTGATTC GTTGATAATG ATCATATTTT	480
	TGTTTTGACC TGATGCACTA AATTGGTGCG TCAGTTTGTC CTCGCTATCA ACATGATTTT	540
20	TAATGGCATT GTCTGATCTT GCATATCCTA AAATCTCAGC AATATCTTTT CCTACAAAAT	600
	AAGGTTGCGT TTCAATTTCT ACTGTTCTTA CTGGTAGCTC TTTAAATTA AATGTTTGTA	660
	ATGCTTGCAT TTGAGTATCC TCCTTTTTC TCAACACCCA CATTGAGCAG ACGGTTATCG	720
25	CAATGACTAT CGAATGTATT TAAACGCGGC TCATATCATC GCCAGCTCTC GCTCACATCT	780
	GCTCAATGTG GATGTTGATA AGCGTGGTTA TATTAAGAAG TGAATGTTAC TGATTCACCT	840
	TCCGCCACTC TGTTAAATCA GTAACCTTGT TATCGCTTTC AACACCGTTA AGCTTGCTTA	900
30	ACGCTTTCAC TACTTTTTGG AACTCTTTGA TAGCACTTCG TAGCTTTTAA GTAATTTTCA	960
	CTTCTACCAT TTCCAAACCA GCAAATGCGT CTTGTTATT CATGCTTAGA TGTTTGTTGA	1020
	AAAGATCTCG AGTGATCTT ATTTCTTTAA GTGATTTATC ATAAGCTTCA ATTTGTCCTG	1080
35	AAAGGTTATG ATATTTTAGT TGTAGTTTTA CTAATTTTAA TGATTGGTCT TGCATTTGTT	1140
	ATGTCTCCTT TAAGATGTTT GTTTCGTTT CGTGACTTT GTGGGTAAAA AAATATCTCC	1200
40	AATATTTTCG TCAAAAAAAT CAGCGATAAT AAACATCTCA TCATTCTTAA ATTGATGCTT	1260
	TCCTAATTCT TTTAAACGAT AACCTTCAGT TGATATATTC AAGAGGTTTG CTAATCTTC	1320
	TTGAGTACAC TTTCTTTCTT TTCTCAACTT TATTAAATTC CATTGCATGT TGTCACCTCC	1380
45	CGCTTTACAA AACCTACTAT ACACGATACG GTACTTGnGT CAACATAAAA GTTTGCTTTT	1440
	CGTGTATTT	1449

## (2) INFORMATION FOR SEQ ID NO: 593:

50

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1170 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 593:

5 ACCAAAAACA CTTTCTAAAG AAATTGTTTC AGAACCATT T GCTAAAAACC AATTACGTGA 60  
 AAAAGCTCGT CTAATAACA TTAAAGGTTT AGAGATTCCA AGAGTTTCAT AACTTTTAGA 120  
 CGATGATGAT TTCATTACAG ACGTAGAAAC AGCAAAAGAA TTAAAAGCAA AAGGTGATAC 180  
 10 AGTCAAGTTC ACTACTAATA AATTCAAAGT ATTTGCTGCA ATTCAGATA CTGTAATTCA 240  
 TGGATCAGAT GTAGATTTAG TAACTGGGT TGAACGCA CTACAATCAG GATTAGCAGC 300  
 TAAAGAGCGT AAAGATGCCT TAGCAGTAAG TCCTAAATCT GGATTAGAAC ACATGTCATT 360  
 15 TTATAATGGA TCTGTTAAAG AAGTTGATGG AGCAGACATG TATGATGCTA TTATTAACGC 420  
 TTTAGCAGAT TTACATGAAG ATTATCGTGA TAACGCAACA ATTTATATGC GATATGCAGA 480  
 TTATGTCAA ATTATTAGTG TTCTTTCAA TGAACAACA AATTTCTTTG ACACACCAGC 540  
 20 AGAAAAAGTA TTTGGCAAAC CAGTAGTATT TACAGATGCA GCAGTTAAAC CTATTGTGGG 600  
 AGATTTCAAT TATTTTGGAA TTAATATGA TGAACAAC TATGACACTG ATAAAGATGT 660  
 25 TAAAAAGGC GAATATTTGT TTGTATTAAC AGCATGGTAT GATCAGCAAC GTACATTAGA 720  
 CAGTGCATT AGAATTGCAA AAGCAAAAGA AAATACAGGt CCATTACCCA GCTAAGCCCC 780  
 AAAAGGTTAA TGTAACAGCT AAGGCTAAAT CAGCTGTAAT ATCAGCCGAA TAGGGGTGAT 840  
 30 GAAATGAGTT TAGAAGAAk TAAATTGTGG TTGAGAATTG ACTATAATTT CGAAAATGAT 900  
 TTAATTGAAG GTCTCATTCA ATCGGCTAAG TCTGAATTAC TATTAAGTGG GGTnCCAGAT 960  
 TATGACAAAG ATGACTTGGA ATACCGCTT TTTTGTACAG CGATTAAATA TATCATTGCA 1020  
 35 AGAGATTATG AAAGTCGTGG ATACTCAAAT GACCAATCTA GAAGCAAGGT GTTTAATGAA 1080  
 AAAGGATTGC AAAAAATGAn TTGAAATTA AAAAAGTGGT AGGTGATTTT TAAATGGAAT 1140  
 40 TTAATGAATT TAAAGATCGC GCGTATnTTT 1170

## (2) INFORMATION FOR SEQ ID NO: 594:

## (i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 1393 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 594:

50 TGAAGGTGT TTTGTTTATG GCCAATTGCT GTGTTATTAA AnCGAATGTT TCGAATGGAA 60

55

ATCATTCCGG AAGAGGACAT GTCTTCAGCT TCTGATTTGA AAGCACAGTT TTATCGCGGT 180  
 ACTTTGCAGC GCTTTTACCA AtCnTTGCAC GCAGAAAAGC TTACACCTTA TGTTATGTCT 240  
 5 TATGACGATA TCATTTCATT TTGTAAAGAA AACAAATATCT CTGAAGTAGT GACTGCGGGT 300  
 GATATTATGA GTTATCATCT TGAAGAATAT GATATTTTAC ATCAACGTTC TTTATTCAAT 360  
 GAAGCACGCA TTGCCGTTAC TTTGATACGT GGGAAATCATT ACTTTAAAGC GAGTAAAACA 420  
 10 ATGAATCAAC AAGGGGAGCC ATACAATGTT TTTACTAGTT TCTATAAAAA ATGGCGACCT 480  
 TACTTGAGGC ATAGAGACGT ATATCACTAT GATTTAAAT CATTGAAAA CTTTGTCAAT 540  
 GCATCACCTG ATGATTTAGT GTTTGATGAC ATAGCATTTG GATCCTCACA aATAATTGAA 600  
 15 CAGatAAATG GCAACATTTT TTAGATCAAG ATATACAGAA TTACGAAAGC GGAAGAGACT 660  
 ATTTACCTGA AGTATTAACA AGTCAGCTAA GTGTTGCTTT AGCATATGGA TTATTAGATA 720  
 20 TTATTGAAAT TTTTAATGAT TTATTGGCGC GTTATGATGA AGATGAGGCA AACTATGAAG 780  
 CATTTATACG TGAACTCATT TTTAGAGAAT TTTATTATGT GTTAATGACA CAGTATCCTG 840  
 AAACCTCATA CCAAGCTTTC AAACCTAAAT ATCGACAGAT AAAATGGTCG CAAAATGAAG 900  
 25 CGGATTTTAA TGCATGGTGC GAAGGGCAAA CAGGATTTCC AATCATTGAT GCAGCAATAA 960  
 TGAATTGAC ACAAACTGGT TTTATGCATA ATCGAATGAG AATGGTTGTG TCGCAATTTT 1020  
 TAACCAAAGA TTTATTTATA GATTGGACAT GGGGAGAAAA ATTCTTTAGA AAGCACCTTA 1080  
 30 TTGACTATGA TGCAGCATCA AATATTCATG GATGGCAATG GTCTGCTTCT ACAGGTACGG 1140  
 ATGCAGTGCC GTATTTTAGA ATGTTTAATC CAATAAGACA GAGTGAACGC TTTGATGCTA 1200  
 AAGCTTTGTA TATCAAAACA TATCTCCGA TTTTAAATCA AATTGATGCA AAATATTTGC 1260  
 35 ATGATACACA ACGCAATGAG TCCAACCTTT TTGAACAGGG GATTGAATTA GGTAGTCATT 1320  
 ATCCAAGACA AATGGTAGAT CATCAAGAAA AACGTACACA AGTTTtagCT ACATTTAAAG 1380  
 40 CGCTAGACTA ATT 1393

(2) INFORMATION FOR SEQ ID NO: 595:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1484 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 595:

TTAACTTGAA AGATTTCCCC GACATCTATA ACTTGTATAA TGTGTATATG TCGTTTTTAG 60

	AAAAATTAAA AAAAGAATAT CTTGATTGGT ATAATCAGAC CTTAGAGTTT TCTAATTTAT	180
	CAACAATGT AGTAAGWATA GATACTCCTT TTAAAGATAA TTCTTTAGAT AATTTAATAA	240
5	TTTACGCTTT ATACGATCAG TCCAGAGACA TGATTACACT GACAGATGAC GGCTATACTA	300
	TATtTGATTT AGAAAATAAT GGTATTTCTT TAAATAAATC AAAAAACGT AAAAAGATTT	360
	TTGAAGAGCA CCTTTCAGCT TACGGTATTA AATATAACGA TAAAACTCAC GAAATTTTTG	420
10	TTCAAATAA CTTTAAAAAT TTTAATAAAT CGAAACATAA TTTATTACAG TGCCTTATAT	480
	TTGTTAATGA TATGTACTTA CTTTcTAATC CTAAGTCACA GAACATATTT ACAGAAGATG	540
	TTGCAAACAA ATTGGATGAA CATAACATTT ATTACGGAAG AGATTTACCT ATTATAGGAA	600
15	GCAGTGGTGT TGTTCATAAT TTCGACTTTT TTATTAGCGC TAAGAAAAAT CAAAAAGAAA	660
	AATTTATCAA TGCTATTTCT AACCCATAA ATTCTATGAT TATTAAGTCG AAAATAACGG	720
20	ATGCTATGCA AGCAAAAAAA ATAAAAAGAC ACAGGCAAAA TGAGTTTATT TTTATTTTAA	780
	ATGACTCAAA AAAAGAAATA AATGAACATA ATAAAAATCT TCTTCATGAA AACTATATTA	840
	GTACAATAGA TTATAGCGAA TTAGATGAAA AGATAGGTTT ATTGATTTAA TATATATAGA	900
25	CGTGATAATA TCAATGTTTA TATTAAATGA AACGAATTTG AAAATTTCGA AACAGCTTAA	960
	GATAGCAAAT TGAATAGCCT TATTGATAAT GCAGAATCGT CTACACTTAG TTGAACAAAT	1020
	TCTATGAGAA TAGATATTGT TAAACTATTT GGGTAGGCGA TTTTATTTTG ACAAGAGTCA	1080
30	GAAGATTATT TAGTTCATAG TTAAAGTTAC AAATGGCTAA ACTTTATGAA AATGGTATGA	1140
	CTAGAATCGA AATTATACTC GAATACGATT TAACACTCTC AATCTTCTCG AACTGAATAA	1200
	ATACTGGGTC ATTCAATCAT CAAGACAACCT TAATAAGTGA TGATAAATGT TCATACCATT	1260
35	TTGCAAATAT AACCAGATAC TCTCAACAAA ATGTAGAGTT CAAAATTAAA AATGTCCCTA	1320
	ATTGGCACAT TGCTGTATGA TAATAAATTC AACTTAGATT TCAACAATAT AACTAATGAT	1380
40	GCTTATATTT ACCGCGATAT CAAACCAAAC AGCCCATCCA CAAACAAATT AGAATATATC	1440
	mTGTTAATAA TAGTAAATGA TATATCTTCC AAAGACCGCA AACT	1484

(2) INFORMATION FOR SEQ ID NO: 596:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1435 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 596:

55

CTTTATCaTT AAATGCCaTT TTGaATCCTT CTTGAATTTG TTTGATTTC TTTTCAGTAT 120  
 TATTAACCTT GTTCAAGACA GTGCCATTAA TTTTCTTCTT GAGTGGACCA ACCTCATCTT 180  
 5 TATTAGTTGA TTCATGAATC GTTAATCCAA TGTGTGGCTG TATTCTATAA CCATCATTCG 240  
 CTATAGTTGA AACATATTGT GaTAATTGTA ATGGTGTATA GGTATCATAT TGACCAATTG 300  
 ATAAATCTAG ATAATTACCT GGATTATTTG TTAATGGTTC GATTTGACCT CTTGTTTCAT 360  
 10 TTGGTAAATC TATCCCTGTT TTCACACCTA AGCCTACTTG ATTTAATCCT CTTCTTAGCT 420  
 TTTGGGCAGG TGAAC TTATG TCTGAAGGTA AAGCCATACC AGAATAATAA GGGTCTCCCC 480  
 15 CTAATTTTAA TGCTGTTTTA AACATATATA CGTTTGATGA ATGCATCAAA GCTTGCTTAT 540  
 CATTAAATAGT TACATGCCCC TTTTATTGA AGTATGATCG TTTTGTCAAA CCACCTTGGA 600  
 AATGTAATGG TTCATCGACC ATTGTTTCTC CAAC TTTGAT AGCTTTATTC TGATAACCGG 660  
 20 CTAATAATGT TCCACCTTTT ACAGAAGATC CAACCGCAAA TTGAGAAGTA AACGTACCAA 720  
 TGTCAATATC AGTCATTTTA CCACTCTTAT TAATCTGCTT TCCGGCAAGC GCAAGAATGT 780  
 CTCCATTTTT AGGATTTTGT ACAACCATCA TTGcATTATC CATATCTTTG GCACCTTGAC 840  
 25 TGCGAsTTCT TAATTTGTTT ATCTAATAAT GCTTCTACTT CTTTTTGAAG ATCTATATCG 900  
 ATCGTTAATT TCAAATCTTG ACCGCGAGCG CCAGGrTTTA ACACTTCTGA AGATGTAAC T 960  
 TTACCAGATT TGTCCGTTGT GTATTTcATT TCTTTCTTCT TACCACGcAA TACATCTTCA 1020  
 30 TATTGATATT CTAGGTAAGA TTTTCCAACA CGATCATTGC GTGAATATCC TTTGGATAAG 1080  
 TAATGTTCTG TCAATTCTTT TGGaATACCT TCAGCAGGTG TCGATACATC TCCGAATATA 1140  
 CCTCTTAAAG TATCGCCATA TGGATATTTT CTATCCCAAT CCATAGACGT GTTAACACCT 1200  
 35 GGTAATTTGG AAAGTTGCTG AGAAACTGCT GCATACTCTT TTCACTGAC ATCTTCATTT 1260  
 TTTATCATTT GTGGATCTAA AACTGTTcCT GGCATTcATC TCTCGAAAAA TAGCTAAAAC 1320  
 40 TTGGTAAATC TTTAGAAGAC AATTCATCTA ATTGTGGATT TTCCGGATTT CGGATAACAG 1380  
 TTGGTnTATC CATAATGGAT CCTGGTTTAA TACCTTCCAn CnGGCGAACA TAGCC 1435

(2) INFORMATION FOR SEQ ID NO: 597:

- 45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 516 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 597:

55



AGAAAAAGAT TTAGAAAAGA TTACCATTCG CGACATAACA ACACGCGCTG ATATCAATAG 120  
 AGGGACATTT TACTTACATT ACGAGGATAA ATATATGTTA CTCGCAGATA TGGAAGATGA 180  
 5 GTATATTTCA GAACTAACGA CATATACTCA ATTTGATTG TTACGCGGTT CTTCArTTGA 240  
 AGACATTGCG AATACTTTTG TAAATAATAT ACTCAAAAAT ATTTTCCAAC ATATTCATGA 300  
 TAATTTAGAG TTTTATCATA CTATTTTACA ATTGGAACGC ACGAGTCAAT TAGAATTGAA 360  
 10 AATCAACGAA CATATTAAAA ATAATATGCA ACGTTACATT AGTATCAATC ATTCTATCGG 420  
 AGGCGTTCCA GAGaTGTATT TTTACAGTTA CGTTTCCGGA GCAACAATTT CaATTAnnTA 480  
 15 AATACTGGGG TAATGGACAA ACAGCCCATT TTCAGT 516

(2) INFORMATION FOR SEQ ID NO: 598:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1955 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 . (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 598:

AGTACTTTAC CATGAGGTAA GACTTTATTC CCTGCTTTTA TAGATTGTGC TTTTATTTGT 60  
 GTGCCAGAAC CAATAGtAct GGTcTAATGA TTTTGCACTC ACATTATCAA TAGCTTTTTG 120  
 30 TTTGTCTTGA CCTTCAACAT CTGGCACTTT ACTATACTCT GCATTAGATG TGTCACTTTT 180  
 TGATTTACCT ACATTTAAAT ATTCAAAGT ATTTTCCATT ATTGGTTTAA ACGCTTTACT 240  
 AACACCTAAT TCATAAGCTT CTTGGTCATT TTTTGTGACC AAGCTCATAC CAGCGTATAC 300  
 35 AATAACTTTA GGATTTTCT TCGGCGCGTC ACCCATAAAA CTTACAAAAT ATGGGTTTGG 360  
 ACCTTTAACG TATCCACCAC CATTAGGTGC AGCGACTTGT GCTGTACCAG TCTTACCTTC 420  
 40 GACCTCATAA CCATCAATAC GATAGTTTGC AGCGTGACTC TTCTTACTAT TCACAACTAA 480  
 ATCCAATTGC TTTTCAACTT TTTCAGCAGT ATCTTTTGT ATTGGTTTGC CTGCGATTG 540  
 TTTTGGCCCT TTATAAAATT GTCTTTTACT AACAGGATTT TCAACGCTAT TCACAAACCA 600  
 45 TGGTTTTAAC ATATTACCAT CATTAAAGAA CGCTGATTGC GCTTGTAACA TTTGAACAGG 660  
 TGTTACTGTT GTCGATTGAC CAAATGATGA CGTTTTTTGt TGCAACTCAT TACTCCATCC 720  
 AATTTGaCCA GGTGCTTCTC CATCAAACAT ACCTTTAGTT GaTTTTCCaA ATCCaAATCG 780  
 50 TTCATACCAA GATTTCAATT TGTCTGCACC AACTAAATCT TGTAATGCA TCATCAATGT 840  
 ATTAGATGAA TAAGTAAATC CGAGTGACAT TGGGATTTCa CCCCACCga CTCTATTCCA 900

55

AAAAGCACCT TCTTGAATAG CAGCTGCTAA CCCATATGAT TTAAATGTTG ATCCAGGCTC 1020  
 GTATGTGTTT TGATAAAGGT CATTTGCCCA CTTTTTACCA AAGTCTTTAC CAGTTTCAGG 1080  
 5 ATTAAATGTT GGTGCTGAC TGTATGCTAA AATTTCTCCA GTTTTGGCAT CCATGACAAC 1140  
 CGCAAATAAA TCTTTGCGCT GGTATCTTTC AACCATGCCA TCTAAAGCTT CTTCAACAAA 1200  
 10 TACTTGAATA TTTGAATCGA TTGTTAAATG GACATCATCA CCACGTTTAG GCTGCTTCTC 1260  
 TTTTTTAGTA TTTGGTGCGA TATATCCCCA AATATCATGA ATATATCTCA ATGATCCTTT 1320  
 AGATCCACTT AAATAACTAT CAAAAATCTT TTCAACTCCA AGTGCACCTT TAAGTTCACC 1380  
 15 AGTATCCGGA TTTTCTGAG CTCTACCAAT TAAGTGTGAT GCAAAATTGC CATTGGGATA 1440  
 AAAGCGTCTT GTTTCAGGCA ATAAAGAAAT ACCAGGCAAA TTCATTTTCT CTATTTTCAA 1500  
 TTTGTCCTGA TACGTTAAAT TTGTTCTTTT GCGTCCAAAT TCAATTTGGA AAGCTTTCTT 1560  
 20 TTGACTAAGT CTCTTTTCAA TTTCTCTGG CTTCATATTA ATGACTGTAG ATAATTCTT 1620  
 TGCAGTCTCT TTTTATCAA CTACATGCCT AGGTTTTTTA GAATTGGCAC TCGCCTTTT 1680  
 ATCTATTACT GCAACAAGTT TATATCTTTC TACATCTTCT GCTAGCACTT TACCATKACG 1740  
 25 ATCATATATC TTTCTCGTT CTGGTTGTTG TGCATTCTTA ACTAAATACT TTTCAATTGC 1800  
 CTTCATGACT AAATCTTGAC CATTAGAATG TCCAGTAATC ATGATATATG AAATCTTAA 1860  
 AACCAATATA AAAAAGAGCA GTCCGAATAA ACCAACAAGT AGGACTGCCC CTATTTTATT 1920  
 30 TTTTTTAATT TTAATTTTTT TGGTCGCCAT TACTA 1955

## (2) INFORMATION FOR SEQ ID NO: 599:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 704 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 599:

TAGCAAGAGT AGGTGATAAC AGTTCTTTTA ACGTGTGCTG TGCTTGCTCA TTTCTGGCT 60  
 45 CATTGTAGAC TAAAATATAA TAAGCGTTTT CAAATAAGTG CTTTTTAGCA TTTAGTACAC 120  
 CAGATTTATG ACTACCAGCC ATTGGATGAC CACTGACTAA ATGAATATTA TGCTTTAATA 180  
 AATTGCATTC GTGTTGCTGT ATCATTGCTT TAGTACTACC AGTATCAGAA ACAATAACAC 240  
 50 CAGGTTTAGT TGGCATATCT ATAAGCTCGC TAAGATATTT ATTTGTGATA GCAACAGGTG 300  
 TTGCATAAAT AATTACATCG GCTTTTTTAA TAGCTTCACT ATAATTTAAA CATTTTTCAT 360

55

TATTAGTATT AGGGTTATGG TATTTTATAT TGCTAGCAAG ACTTCCACCA ATTAATCCAA 480  
 GCCCAACAAA TAAAACTGTT GTCATATAAA TCACCTTATT TCGAAATTTT CAGAATAATA 540  
 5 ACATTGTAAA TGAGCTGTTG ACACAGTGCA ATAGTAAATA AAAATCGATA ATAGCATTAA 600  
 TAGATAAACG GAGATAAATC ATCTACAATA AAGAGTATAG TAACACAATG GCAACGGAGG 660  
 GGTAATCAA TGGAACCAAT ATTAGAAATG ATTAAAACAT TAAC 704

10

(2) INFORMATION FOR SEQ ID NO: 600:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1158 base pairs  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 600:

CCTCTAAAGG ATCACGAACC TTTTTCATCG TGACTAAGCC TATAATACCC TTAAAcCTAT 60  
 TATTATTAAC TTTTACCTCT GTGTATTCCC TATCAATTAG GCGACGCCAG TGACGTTTAT 120  
 25 CTATATACTT TACTTTCACA GTCACCAACT CCTTGTCAAT ATTATATAAA ACTACATAAA 180  
 TGATGTCATG TCATGATACA GTTTTATAGT TTTTGTTAAT CATATGAAAA TATTTATTAT 240  
 TATTTTACTA CAACTCGCTT CAATTTACTT AAAATAGACA ATATTAATTA GATAGTACAC 300  
 30 ACATTTCTTC ATAAAAGTGA TTTTCAAAA ATATAAATAA CACACTCTTA TCGTTTTCAA 360  
 AATCATTTAA TGCTATTTTC ATTAAAAACA GCTGAAGCAT CAAATCTATT CTGATTCAAT 420  
 CAAGAATACA TATAAAAATG AAGTGACTCA AAGGTTTATT AACAACTTC AAACCACTTC 480  
 35 ATTGATCACT TTTATTTTAA AGCATATTTT TCGATTACTG ATTTAAGATG CGGATATTGC 540  
 GTAATTAATT CAGATTGCTT AAACAATTCA AACTGCTTAA ACTCAAATGC CGGTTGACAC 600  
 40 ATACAACCTA CTAAACTAAA AGTATTTGAT ATTTGATG AAGAAGCAAA AATTGTTCTT 660  
 TTAGGCACTA CATATTGCAA TACATCTCCA TTTTGGATAT CAGTACCCAA TGTTCAGTC 720  
 GTATATTCCC CATCCGGATT TATCATATGA ATTGTTAGAG AATCGCCAGC ATGATAGTAC 780  
 45 CATACTTCAT CAGCATCAAT TCGATGAAAA TCGGAAATAT TGTCATCTGT AAGTAAAAAA 840  
 TAAATACTAC TAAACGGCGC TCTGCGTCCA TCTTTCAATA CTTCTCGAAT TGTCTCTCTA 900  
 TAGAAACCAC CTTCAGGATG TGATTCAAGT TGCAATTcAT CaATCCATTG TTCTGCTGAT 960  
 50 TTCATTATTT CAAATCCACA TTATGGAATA CGTTTTGTAC ATCTTCTAAA TCTTCTAATG 1020  
 CATCGATTAA TTTTTCAAAT GTTACTTGGT CCGCTTCAGA AAGTTCAATA TCTGTTTGAG 1080

55

GAACGACTGC AAATTGAT

1158

(2) INFORMATION FOR SEQ ID NO: 601:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 827 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 601:

15 TGATACTTAT TTTTCAATAT TATTTGGAAT AAAnTCTTnT AATTGTGGTG TGAAATAAAC 60  
 ACCGAAATTC TTTTCAGATT CATCTTTTcC ACTACCTGCA TATAAAATAC CAATCAGTyC 120  
 ATGTTTAGAA TTCAATACAG GAGATCCTGA ATTACCTGGT TGTGCATACG CATCAAATTC 180  
 20 CATAAACGTT CCACTGATAT GGTTAATCGT TCCTGTCGAT TCAAACATTT TATATTTTGT 240  
 TTGTGCACCC TTTGGATAAC CAATAACAGA AATTCTATCT TTCACTTTTG CTCCGTCTGC 300  
 AAATTTTGTA TAACTAACGT TCTTATTAAA ATTCAAACCT TCTGTACTTG TTTCATGAAC 360  
 25 ATGAACTATC GCAAGGTCTT CTTTTCCGGG ATATTCTACA ATGTCTTTAA CGTCGTAGTT 420  
 TCCTCCGCCT TTACCTTTAC TCGAATGATG TGCTGATACT CTATTTTTTAA AAATATCATT 480  
 ACTTTTAGCG ATATGTTTGT TAGTTACGAT TGTATTTTAA CCAACAACATA CACCAGTACC 540  
 30 ACCCACAAAT GCTACCACTG aATTGTATGG TTCCTTAGTT GCATCGGTAA TTTCTTTGAC 600  
 ATTCCTTTTCT GCTTTGGCAA TTGAATGaRg CTGATTAGAA ATATTTTCAG CAAAACCAAG 660  
 AGATGTTAAA ATAGTTAAAG CAGTTAAACC TTTAACCATT ACATTTTTAT TCATATTTTA 720  
 35 CCTCCTTCCA AATTTATTTT nATAGTTAAT GTATCTAGAT ACACATTTAC TATATACATT 780  
 TTAAATAAAA ATATTnATTA AATATTAATT AACCACCTAT nAACGTT 827

(2) INFORMATION FOR SEQ ID NO: 602:

40

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 796 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 602:

50 nGGnACGTAA AGTATTTATG TTAAATGGCA nAAATATATG AATGATGAGT TACCTGGATT 60  
 ACCAATGTTC CAAGGTAAAT CGATAACTAT TGTTAACGAT AAAGTACGAA ACTTAGACAT 120

55

CGACGATTTT GTCGAGGTTT ACTTATATGG GGGGATGTTA TGAATAATGT ATTGTTAGAG 240  
 GTTAAAGATT TAGAAACATC ATTAATAATA AATAATGAAT GGTTAGCAAC TGTGAAAAT 300  
 5 ATTTCTTTTG AATTATCTAA AGGAGAAGTT TTGGGTATAG TAGGGGAATC TGGTTGCGGT 360  
 AAGTCCATAT TAAGTAAGTC AATTATTAAA TTATTACCAG AAAAGATATC TAAACTAAGT 420  
 10 AATGGAGAAG TTATATTTGA TGGTAAACGA ATCGATACGC TCAATGAGAA GCAATTGCTA 480  
 GATATTCGAG GAAATGATAT TGCTATGATT TTTCAAGAAC CTATGACTGC TTAAATCCT 540  
 GTATTTACCA TAAAAATCA ACTTGTGGAA TCTATAAAAT CACATAAAAA AATTTCTAAA 600  
 15 AAAGAAGCAA ATAAATTAGC AAAAGATTTA CTAAAAAAG TTGGAATTGC TAGACAAGAT 660  
 GAAATATTAA ATAGCTATCC TCATCAATTA TCTGGTGGTA TGAGACAAAG AGTAATGATT 720  
 GCAATGGCCA TTTCATGTTT TCcTAAATTA TTAATTGCTG aTGAACCTAC AACAGCATTG 780  
 20 GATGTCACGA TTCAAG 796

## (2) INFORMATION FOR SEQ ID NO: 603:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 1618 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear  
 30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 603:

GGATGTGTTT ATCCCAAATT GaAaCagAaA TTTATGATCA AAtGATGTGG ATTGGTTTTA 60  
 35 AaCCATTCAa AATTaCCAAT ATTAmCAAG AATCTGAAGA CATTAAATCA TTTtACAGTT 120  
 GAAACTGAAG AATATGACTT TAGTGAATTT ACACCAGGCC AATACATCAC AGTTGATGTT 180  
 40 TCTAGTGATA AACTTCCATA TAGAGCTAAA CGTCACTATT CTATCGTATC AGGTGAAAAA 240  
 AACCATTTAa CTTTTGGCGT TAAACGTGAT GTCACAACAG AACATGAAGG CGAAGTTTCA 300  
 ACAATTTTAC ACGATGAGAT TAAAGAAGGC GATATGATTA ATTTAsTGCG CCTGTAGGTG 360  
 45 GATTCTGATT AGAGAATACG ACTGAACCAC AACTTTTCTT AGGTTcAGGT ATTGGTGTTA 420  
 CACCTTTAGT AGCTATGTAT GAAGCTGCCT CTGCCAAAGG TTTAGATACA CAGATGGTTC 480  
 AAGTTGCTGA AAATGAACAA CATTTACCTT TCAAAGACAA CTTCAACAGT ATCGCAAGCC 540  
 50 ATCATGACAA CGCTAAATTA TATACACACT TAnAgATAAA CAAGGCTATA TTGGTGCTGA 600  
 AGAATTACAA GTATTTTtAG CAAATAAACC TGAAATTTAT ATCTGTGGTG GTACAAAATT 660  
 55 CTTACAaTCT ATGATCGAAG CACTTAAATC TTtAAATTAC GATATGGATC GCGTACACTA 720

ATATGAAACT TAATCAACGT TACGTAAAAG TATTGTCATT ATATTTCGTA AGTATTGTTA 840  
 CTGCAAATAT TATTGTTAAA AATAATAATT TAATTAAAAC AtTGATACAA ACCATAGCCG 900  
 5 GGTACACGGT CTTTGCAGTT GGTTTGAAGT ATTAACTAA ACGTAAAAAT AAATGACATC 960  
 TATACCAAAA CAGCTATGAC TTTACTTTGT CATGGCTGTT TTTTAATTGG GAGTAGGACA 1020  
 GAGATGATAT TTTTCGCAAAA TTTATTTGGT CGTCCCACCA CAACATGCAT TGATGTATGC 1080  
 10 TCACTGAATT TCATAAGAAA GGAGTTCACA AGATGACCGT AGATATTGGA CGGATTTATG 1140  
 ACAATAAAGA TAATACCGAC GCTATTCGTA TCCTAGTCGA TAGAGTCTGG CCGAGAGGTA 1200  
 15 TTTTCGAAAAG AACTGCTAAC CTAGATTATT GGTTAAAAGA CATTGCCCCT TCTACTGAGT 1260  
 TGCACAAATG GTTCCAACAT GATCCTAAAC TTTTGGGAGC TTTTAAAGAA AAATATGAAA 1320  
 AAGAATTACG TGATCAGGAT GCGCAAAAAG ATGCTTTTGA AAAATTAAAG GATATTGTAA 1380  
 20 ATCAGCATAA TCATGTTCTA TTGTTATATG CAGCAAAAAGA TACTAAACAT AACCAAGCTG 1440  
 TAGTACTACA GCAGTTGCTC AATACTTAGT TATATATGTT TATCGTGAAT CACTATAAAT 1500  
 GTCGCTAACT TCATTTCTAA TATTTATTTT TAAAAAGCAT AATTACTACA ATTAATTGGA 1560  
 25 ACTTTAATAA TTAATAAAT TGGAACAATA TTTTACTTTA ACAAATAAAT GTTTTAAA 1618

## (2) INFORMATION FOR SEQ ID NO: 604:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2115 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 604:

40 ATTATAATGC AAATATTGAA GCATTTAAAA CAGTCGCTAA AGCAGTAGGC AAAGAGAAAG 60  
 AAGGCGAGAA GCGTCTGGAA AAGCATGATA AAATATTAGC GGAGATTAGA AAGAAAATTG 120  
 AACAGAGTAC GTTAAATCT GCATTTGCAT TCGGTATCTC AAGAGCAGGT ATGTTTATTA 180  
 45 ATAATGAAGA TACATTTATG GGACAATTCT TAATTAAAT GGGTATTCAA CCTGAAGtca 240  
 mAAAArAmAA AACTACGCAT GTTGGTGAAC GCAAGGGTGG TCCTTATATA TATTTAAATA 300  
 ATGAAGAACT TGCCAATATC AATCCAAAAG TTATGATTTT AGCCACTGAC GGAAAAACGG 360  
 50 ACAAAAATAG AACGAAATTC ATTGATCCTG CAGTTTGGAA ATCATTAAAA GCTGTGAAAG 420  
 ATAACAAAGT TTATGACGTT GACCGAAATA AGTGGTTGAA ATCAAGGGGG ATTATCGCAA 480  
 55 GTGAAAGTAT GGCAGAAGAT TTAGAAAAAA TTGCAGAAAA AGCAAAATAA AAATACAGCG 540

TGTGATGATT TACCTGAAAG TTTATTTCTGA ATAAATTTAA TTACATAACC GACAAGGATT 660  
 GTTTTAACAG TTCTTTTAAT GAATTGGCGC ATCGTTACAT ACCTCATTTT TCTATATCTT 720  
 5 ACGAACTATA TACCCATTCA TATATGCTTT TTAAACGTCA TTGTCACAAT TTAATTTTTTA 780  
 GCGAATATAA TATAACCATC TTTATCTGCT TTTTITAGTAA AAATGACAAA AATTGCATGT 840  
 ATTATTGaGA TGATGGTAGG GATACCTGTC CAGAAAAATA ATAAGTGAAA AAGACCTTGT 900  
 10 CCAAATTTAT CAGCATAAAA TTTATGaATA CCTAACCTC CAAGAAATAA TGCAACmATa 960  
 ACATAAATGG CTTTATTGAC TTTCAATTGT AAtCCTCCTT AACTATAATT CTACTTAAAT 1020  
 15 TCGTTGTGAA AACCAATATT TCTAACTTGA GAATTTTCAA ACTTTCTAAA ATTATAAGTA 1080  
 TATCTTTTTA AAATAAGCTA GAATTTCTAT ATAATAAATG TTAATAACGT AAAAGGGAAT 1140  
 GATGACATAG TGATACGTCA AGCACGTCCA GAGGACCGAT TTGATATTGC GAAGTTAGTT 1200  
 20 TATATGGTTT GGGATGATAT GGAATTAGAA TTGGTAAAGC ATCTACCTAA AGACATGGTA 1260  
 TTAGATGCAA TTGAAAAAAG CTGTGTTGAT GCAACATATC GAACTTTTTA TCAGCATATT 1320  
 TTAGTTTATG AAGTAGAAAA TAAAGTAGCA GGTGTGATTA TTAGCTATAG TGGTGAAAAAT 1380  
 25 GAATTGAAAT ACGAAAAAGC ATGGGAACTA CTTGACTTGC CAGAAAAAAT AAAACAATAT 1440  
 GGCACGCCAT TACCTGTAAA AGAAGCTAAA GACGATGAGT ATTATATAGA AACAAATTGCG 1500  
 30 ACATTTGCAG CATATAGAGG TAGAGGCATC GCGACAAAGT TATTAACGTC ATTACTTGAA 1560  
 TCAAATACAC ATGTTAAATG GAGTTTGAAT TCGGATATTA ATAATGAAGC AGCATTAAAG 1620  
 TTATATAAAA AAGTAGGCTT TATATCTGAT GGACAGATTG AATTATACAA GCACATGTAT 1680  
 35 CATCATTTAA TTGTTAAATA AAATACTCGA CAGTTCGATG TAAGTCGATT GCCGAGTAGT 1740  
 ATCATTTTCT ATTAAATGCC TGCAAATAAT GCACTAATAT AAATACCTAA TGCATATAAT 1800  
 AAACCGAAAA ATGTATTTGT TTTACCAGCA GCAGCCATTG CTGGCATCAT TGTAGGCGGT 1860  
 40 GTATCATTCT TCTTGAAACG TCTGATAACT TTAACAGGCA TTGGGAATGA TAACAACGCA 1920  
 AGTAAGTAAA ATAATGAGCC ACCAGGTTTA ATAATGATCG TAAGTACAAT AAAGGCATAA 1980  
 45 GCGATAAAGT ACATGATTGC CATAAATGTT AAAGAAGCAT TTTTACCTAA TAGAATGGGT 2040  
 AAAGTTTTGC GACCACTTGC TTTATCTTTG ACACGGTCGC GAATATTGTT AGCCATATTA 2100  
 ATTAAACCGA TAGTG 2115

50 (2) INFORMATION FOR SEQ ID NO: 605:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1003 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 605:

5	TATGTCTCAA ACTCAAATTA ATCAAATGTT CCAGCAGAAA AACATGTCTA CTGAATTAAA	60
	ACGTCGTTAT GCACAACGTT TATTACAGTT TCCACATGTA CACAATAAAG AATACTTGAA	120
	ATCTTATGCT AAAAACCCCTA AAGAAACTAA AGATAGTTAT ATTTCTGGTT TTAAAGAGAA	180
10	TCAATTGATT AAAATAGAAG CGATTAAATC ATTGTTTGCA ATGGATAAAT CTCCATTAGA	240
	ACATGTTAAA CCTGCTACAA AACCAGACGC TTCTTGGGAT GAGATGAAAC AAAAAGCAGT	300
	TGAAATTGGT AAAGCTGATA CTACATCGAA TAAATTTGGT ATTAGAGATC AATACTGGAA	360
15	ATTAATTCAA GAAAGTAAGC GTAAAGTTAG ACGTGACTAC GAATTCAATG TTAATTCTCC	420
	AGAATTCCAA GATTAGAAT TACTTGTAAG AACAAATGCGT GCTGCTGGTG CAGATGTTCA	480
20	ATATGTAAGT ATTCCATCAA ACGGTGTATG GTATGACCAC ATTGGTATCG ATAAAGAACG	540
	TCGTCAAGCA GTTTATAAAA AAATCCATTC TACTGTTGTA GATAATGGTG GTAAAATTTA	600
	CGATATGACT GATAAAGATT ATGAAAAATA TGTTATCAGT GATGCCGTAC ACATCGGTTG	660
25	GAAaGGTtGG GTTTATATGG ATGAGCmAAT TGCgAAACAT ATGAAAGGTG AACCACAACC	720
	TGAAGTAGAT AAACCTAAAA nTTAAAATAC AAATAGCACA TAACTCAACG ATTTTGATTG	780
	AGCGTATGTG CTATTTTTAT ATTTTAAATT TCATAGAATA GAATAGTAAT ATGTGCTTGG	840
30	ATATGTGGCA ATAATAAAAT AATTAATCAG ATAAaTAGTA TAAAATAACT TTCCCATCAG	900
	TCCAATTTGA CAGCGAAAAA AGACAGGTAA TAACTGATTA TAAATAATTC AGTATTCCTG	960
35	TCTTTGTTGT TATTCATAAT ATGTTCTGTT AACTTAATAT CTT	1003

## (2) INFORMATION FOR SEQ ID NO: 606:

## (i) SEQUENCE CHARACTERISTICS:

- |    |                             |
|----|-----------------------------|
| 40 | (A) LENGTH: 1097 base pairs |
|    | (B) TYPE: nucleic acid      |
|    | (C) STRANDEDNESS: double    |
|    | (D) TOPOLOGY: linear        |

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 606:

	GCTnTTATAT TTAAAAAATT TATTnGCCGA TAATTACCTT TATTAAATCC CACCCATTAG	60
50	GAwTaCGAGT AATAGGACTA ATACTGGAAT CACATAATGT AACATAACGT CCCTCCTTTA	120
	ACTTAATTTT AATTGTAATC AAATTTGACA ATAAGTCAAA mCATTAAATAC CTATGATArG	180
	TATCATTTAT TAACATATGT ATCATATTTT TAATCTTGCG TAATTTTTAT CGTTAACTAT	240

55



CTTTTGTCAT TTTAATTATG TTAAGATAAT AGTAGTTATA GAAGTTCAAT CTATAGGAGG 360  
 CATAGCATGG ATATTCCAAA AATCACGACA TTTTAAATGT TTAATAACCA AGCTGAAGAA 420  
 5 GCTGTTAAAC TATACACAAG CTTATTTGAA GATAGTGAGA TTATAACAAT GGCTAAGTAT 480  
 GGTGAAAATG GACCTGGTGA TCCCGGGACT GTACAACACT CAATATTTAC ATTAAATGGA 540  
 10 CAAGTATTCA TGGCGATTGA TGCTAATAGT GGCACAGAAT TACCAATGAA TCCTGCGATT 600  
 TCATTATTTG TTACAGTAAA AGATACTATT GAAATGGAAC GACTATTTAA TGGATTAAAA 660  
 GATGAAGGTG CCATTTTAAAT GCCAAAAACG AATATGCCAC CATAAGAGA GTTTGCTTGG 720  
 15 GTTCAAGATA AGTTTGGAGT AAGTTTTCAA TTAGCATTAC CTGaGTAAAA GGATTGCGAC 780  
 AGCtTGgAAT GATAAAGAAA CACTTTTTCT TATGCATGCG TTTTACCTAT GTATTTCTAT 840  
 TTTAAGTACA CATTAGCATT TTAGTTTCGC TCATTTTTAA AATCACAAAG ATAATTTGTA 900  
 20 ACTTAAATGA TGATACGTTA TTTAAAAACA CGATACTTCG TTTCAATGAA CGCATTAAAT 960  
 AATAAATAAA CACCTCACCA TAAGAAAGGA CTACTTTCTT TTGTGAGGTG TTTatTTGTC 1020  
 25 GACAACTTGA TTATGATTGT TTCATTTTTT GAATAAGTTC ATAATCAGGT GTAGCATACA 1080  
 AAGTTTTTTG ATTGTCA 1097

## (2) INFORMATION FOR SEQ ID NO: 607:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2031 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 607:

40 GAGGTCTAGC ATGTCTCGTT CAAAAAATA CTTTACTTA TCTAGCTTAA TGATTATTTT 60  
 AAGCTTTTTC TTTAATACAA ATAACGTTTT CCTAAGTGGA CTTTTTAATT CTTTATTAA 120  
 ATTAATACTT TTCTGCAGTG TTATTAAGTC AATTGTACTA ATTTTGTCTA TAATTTTTGC 180  
 45 AGATCGTTCA ATTAAATCAC TAAAGCCTGA TGCAGATTGG ATTAGAATTG CGAGTAAAG 240  
 TTTGCCTTGG aTTATTCTAA TTGTTATTTT AGTACATATC TTTTCAATTG TTCGTACATT 300  
 CGGTTTTATT TAAAAAGTT AATATGTCAT TGTAGCCTTA ATACAAAACA ATACAATGTA 360  
 50 TCATGCTATA ATGAGTAAAA CAATTGATA ACGTTGTTGC GTATAAAAAA TATTAGATTT 420  
 TCGAAATCAT AACTATGCAT CTAATCGCTA TAGTTATACA ACAAGATATA ACATATAATG 480  
 55 AGGTTTGATA ATGCATCGAC AATTTTTGTC GTCGCGTTGC CAAAACCTCT TTTTAAATT 540

	TTTAGGCGCT TTATTACGTT ACCTCATTTT TTTTCTGAAT ACTGACGGAG GTTTTCCAAT	660
	CGGAACACTG ATAGCCAATT TGA CTGGTGC CTTTGTAATG GGATTGCTAA CAGCCTTAAC	720
5	AATTGCATTT TTTTCAAACC ATCCGACCCT AAAAAAAGCT ATTACGACTG GTTTTCTTGG	780
	TGCTTTAACG ACTTTTTTCAA CATTTCAATT AGAATTAATA CATATGTTTG ATCATCAACA	840
10	ATTTATAACT TTACTACTAT ATGCTGTAAC AAGTTATGTC TTTGGTATTT TGTATGTTA	900
	CGTCGGTATA AAAC TAGGTG GTGGTTTATC ATGATATCAA TCATTTTAGT CATGATTGGC	960
	GGCGGTTTTG GCGCAATTGC TAGAAGTGCC ATTACTGATT ATTTTAATCA TAAATTTACT	1020
15	TCAAAGTTAC CTATCGCAAC ATTGATAGTA AATCTAGTTG GTAGTTTTTT AATTGGATTA	1080
	ACTATAGGCT TATCAATTTT AATCTCATGG TTCCCTGCGT TCTTGTTAC CGGTTTTTTA	1140
	GGTGGCTTAA CAACTTTTCT AACGTTAGCC AAAGAACTTA CACTAATGAT GACGCCAAAA	1200
20	TTTAATATTA ACCTTTTTTCT CAATTATTCA CTTTACAAT TCATCATTGG ATTTATAGCT	1260
	TGTTATATTG GCTATCATAT TAAAAAATAA AATGCTTCAT TCAGCAAATA GGTA AATTAC	1320
25	GACACCTTCC TGAACGAAGC ATTTTTTAAT TTTCATGCAA ATTTTAAAGC ACCATATAAT	1380
	GCCTACCAA TTTCAATAAT CTTTGTTGCC GTTTAAATAA TGTGAATGTC AATAAATTCT	1440
	CCAAACTAGT CGAAAAATAA GGGAGTGGGA CATAAATCCC TAAAAAACA GCAGTAAGAT	1500
30	AATTTTCAAT TAGAAAATAT CTTACTGCTG TTCTCTATTT ATACAATACT TCGTATTGAA	1560
	TGGcTTCGCT TTCCTAGGGT GCCGTCTCAG CCTCGGcTTC GACTGGCACT GCTCCCTCAG	1620
	GAGTCTCGCC ATTAATAcTA CGTATTAACA TGTAATTTTA CTTTACATA CTTTAAAAAA	1680
35	TAAGACACTT TGCCCAACTT AACTACCAA TAGAAACCTC TGTTAGAATT CCTCAAAATG	1740
	ATATTTTCGCG ATATGTTAAT GAAATTGTTG AAACGATACC TGATAGCGAA TTCGATGAAT	1800
40	TCAGACATCA TCGTGGCGCA ACATCCTATC ATCCAAAAAT GATGTTAAAA ATCATCTTAT	1860
	ATGCATATAC TCAATCTGTA TTTTCTGGTC GAAGAATAGA GAAATTACTT CATGACAGTA	1920
	TTCGAATGAT GTGGTTAGCT CAAAATCAAA CACCTTCTTA TAAACTATT AATCGTTTTA	1980
45	GAGTGAATCC TAATACTGAT GCGTTAATTG AATCTTTATT TATTCAGTTC C	2031

(2) INFORMATION FOR SEQ ID NO: 608:

(i) SEQUENCE CHARACTERISTICS:

- |    |                            |
|----|----------------------------|
| 50 | (A) LENGTH: 687 base pairs |
|    | (B) TYPE: nucleic acid     |
|    | (C) STRANDEDNESS: double   |
|    | (D) TOPOLOGY: linear       |

55

TCCCCGGGAA TCGAACCCGC GAnCTCCTGC GTGACATGCA GCGGTGTTAA CCGCTACACT 60  
 ACGAGACCTA TTAnATTAAA AACTATGTAT TCGCGGAGGC GGATTTGAAC CACCGACCTT 120  
 5 CGGGTTATGA GCCCCGACGAG CTACCGAACT GCTCCATCCC GCGCTAATAT TATTTTGAAT 180  
 TACCTAATTA ATATACCATA ATCAAAAACC TAAAGTsrG AACTTTTTGA ATTTAATTTA 240  
 AATGTTATCT CTAAATAAT TACTTAAATA TCGTAGCAAC ATGTTCTCTG TTGAACACAA 300  
 10 ATATTAGTAT ATTCATTTTT GTAGTGACG TCAACGACAT TTTCAAAGTT TTTTGTGTAA 360  
 AAAAACGCTT CTTATTCCCT TTTATCATAT AAGTGTCTAA TAGTTGTCAT AAATAGTGkT 420  
 15 AAAGCATTTA AAAAGGTATA GGAGTTATAA AGTTTACAAC GCCTATACCT TCTGAAAAAG 480  
 AAATTATAAA AGCTTGTTAC ACCGCATATT CTTTCAGTCA GCGACTACCA ATATAACATT 540  
 GTAGCCctAA GACATTGCTT GACGCCTCAn TTACAACAAT TTTTCAAAT CAGCAGCTAC 600  
 20 CTACTGACAC AACATAACAC AACCcNTACA CTAnCTATCG TGTCATGTAA TCTTGCATCC 660  
 GATCTTGCAA CGCTGTAAAT GTTTCGA 687

25 (2) INFORMATION FOR SEQ ID NO: 609:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 843 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 609:  
 35 GTGTTGAATA CTTTGACAAA ATGAAATCAT TTGTTCTTTC GTTTCAAATT TAACTGTTTG 60  
 AATTAAATCA GTTCGTTTTT CGTAGTACTT CGGCGTTGTG TTCATATTCA TTTTTTCTAA 120  
 40 AAATAAACTA GTAAACAATG CACCTTTAAG ACTCTGACTG ACAACGTGTG GTGCTAAAAA 180  
 GAAACCTTGA TACATTTCAA GCAATGCATT TAATGATGCA CCCGCTTCTT TACCAATACC 240  
 AGGTGCTGTC AATCTATAAC CACATCGTTC AATTAAATCT TTTCTACCAG CAATGTATCC 300  
 45 ACCAATCTTA GCTAAACCAC CGCCAGGGTT TTAAATTAAT GATCCTGCTA TTAAATCGGC 360  
 ACCACATTCT ATAGGTTTAC GTCTTTCAAC AAATCCCCCA TAACAGTTAT CCACAAATAT 420  
 TAAATATTA GGATGCACGT TTTTCAACCT AGTAATTACC TTTTCAATTT CATCTAGCGG 480  
 50 AATTGAAGGT CTTTGATCAT AGCCTTTTCA ACGTTGAATC GCTATTACTT TGGTGCCTC 540  
 AGAAACCCCa TCTAACACAC TTTCAATATC GATCTTACCT TCTTTAAGTG CAATATCTTT 600  
 55 ATACGATmCG CCATGCTCCA TTAAACTTTT AATACCATTT CCGTTTACGC CAATGACTTC 660

ACTTTGTAAT GCAATAGTAA TCGCATGCGT ACCTGAAATA ATTTGCGGAC GAACAATTGC 780  
 ATCTTCTGCT TTAAATGCCT GCGCATATAT TTCTTAATTG AACGAATTTA GAACTAATCG 840  
 5 TAA 843

(2) INFORMATION FOR SEQ ID NO: 610:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 775 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 610:

GATTATTATA AACCGAATGA AACAAAAAAG TTTTACATA TTTTCAACAA CTTTATTTC 60  
 20 ATAAAGCATC AAGTTGATCT AAAGTTTGAT TCATACCTTG TTCAACACCC ATGTTTATGA 120  
 CTTGTTGAGC GGCTTCTTTT GTTGGAAATA CCGATGTTGA TGTCACCTGTC GTTTTCGTTG 180  
 TATTGCTCTT GGAAAATGAC AAAGTAATTT TCATGCTTGG CATTTTTGTA TCTTTTGCAC 240  
 25 CTTGAGGTGT CGCAAATGAG TCAATATATT CAATTAAATA TGGACGCTTA ACCGTTTTAT 300  
 ATTCTGCTAA TGTATAGCTG GTCATTGTAG GTGTTTTAAT CGCATAAAAT GCATCACCGC 360  
 30 CTGAAACAGC ATTAAAGCGA AACACTTTGG TACTAGCGTC TTTTGGATGA AACCATTTTT 420  
 CAAATAATGC TTTkGTCGTA TAGGCATCAA ACACCTTTTC AATCGGTGCT TCTATCGTTC 480  
 TTGAAAAAAT AATTTTATTA TCTTCAACTT TAATCGTCAT CTTCCCACTC CCTATCTTTT 540  
 35 ATATTTACTC ATCTTAACAT GCATTGAGTA ACATATGTTA CCAAATCATT TGTTATAAAC 600  
 TATAGTAATA TTAATCATTC CCTCTGThGA AATTGATAAC TACGAAATAA AAAACACACT 660  
 CTATTCAGCT ATTAACAGTT GAGTAGAGTG TGCAAACCTG TTGAATTAAT GCGCTAACAT 720  
 40 TTCTTCTTTG ATTTGGTCTT nATnTAATTT TGAAGGATAA TATGTTGGCC AGTTA 775

(2) INFORMATION FOR SEQ ID NO: 611:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 728 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 611:

ATCAAAATTT GATACTAAAG GGCTCACGCT ATTTTATGTA TTTATTGGAT TGATTATGTT 60

TGTGGCAATG TGTTTATTTA AAGTAGAAAA ACATGTTTCT TCACCATTTT TACCTGTGGT 180  
 TGAATTTAAT CGTTCGATTA CTTTAGTTTT TATAACTGAC CTTCTAACAG CTATTTGTTT 240  
 5 AATGGGATTC AATTTATATA TTCCAGTCTA CCTTCAAGAA CAACTAGGAT TATCTCCATT 300  
 GCAAAGTGGA TTGGTTATTT TTCCTTTATC TGTAGCTTGG ATTACATTGA ATTTTAATTT 360  
 10 ACaTCGAATT GAAGCAAAAC TATCAAGGAA AGTTATTTAC TTACTATCAT TTACATTGCT 420  
 ACTAGTAAGT AGTATTATCA TTTCATTGGG TATTAAATTG CCGGTACTTA TAGCATTGTG 480  
 GTTAATTTTG GCAGGATTAA GTTTTGGATA TATTTATACG AAAGATAGTG TGATTGTCCA 540  
 15 AGAGGAAACT AGCCCATTTAC AAATGAAGAA AATGATGTCA TTTTATGGAT TAACTAAAAA 600  
 TCTTGGGGCA TCAATAGGTT CAACAATAAT GGGATATCTT TATGCGATAC AATCAGGAAT 660  
 CTTTGGTCCA AACTTACACA nTGtGTTAAG TGCTGTTGCT GTAATTAGCA TGGGCTTATn 720  
 20 GTTTTATG 728

## (2) INFORMATION FOR SEQ ID NO: 612:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 913 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 612:

TATCAGAGCA AATAAGCAAG TTAGATAAAA AGATTTCTGA CTTACAATTA ATTAGACGCT 60  
 35 CTGTATGTGA ATTTATTAAA GGACTCTCTC TAATAGATAC CAGCATTTTA AACAAAGACAC 120  
 TACAGTCACA ATATGATAAA GAAGCATCTA TAAATATGG TCATACGAAA GCATATCAnT 180  
 40 CATTTATTAG ACGTAAAGAC AGCTTACAAT CGCAGGATAT CAGACATAAA TTGACAACTA 240  
 TCTTCAATAA ATTTAATCAT ATGTCTTTGA GTCATTATCC AATCCAAGAT TGTAGTGATC 300  
 TCGTATTTGA GTGGAAGGCA TTTATGAACA CTATCGCTGA TTTTGmTGAT GAAACATTAT 360  
 45 GCTGTATTGC TAAACATAT GAAGATGATA CGCGTTTCAA AGATTACTTT AATTCATATG 420  
 ATAATCAAAA TTTAGCATCA TACATTTTCAG AAGCTGTTAA TTATTTTTTG AGCAATGTGA 480  
 ATAAGAGCGA CAATTTTTTA TCCTCATACA GATGCTACAA CACTGAATAC ACCAACTAAA 540  
 50 ATAATTGTTC AAAGTAGAAA ACTAATCATT TATTCTCTAA ATCAATAACT ATATTAAAAG 600  
 TTATACCTTT GCAAAGCGAA TTAGTATAGG TTACCGAAAG GAGAAAGGAT TAGGTTCCAT 660  
 55 TCGATTTATA AAAAAATATA TTTGAAACA TACAAAAGCG CCAGAAGATA ATCATTTCTT 720

TCAATGCATA TTCAACTAGC TCAGGTTTAC TCTTTAAGCC AAGCTTTGTC ATAATATGCG 840  
TCTTATGTGC TTCTACTGTT TTCACAGATA CAAATAATTT CTCTGCAATT tCTTTATTCC 900  
5 CGTAACCTTT GGC 913

## (2) INFORMATION FOR SEQ ID NO: 613:

- (i) SEQUENCE CHARACTERISTICS:  
10 (A) LENGTH: 654 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 613:

ACAAGTGTG ATGTCATAGT AGCGTCAAWT GTGCTTTATT TCTGGGACAC ATTTAAATTT 60  
20 TTCCACCCCA TTAATTGGAG CTTAATCTTT ATTACAATTT TACTATTATT AAACATTTTT 120  
TCTGTAAAAT CATTTGGAGA AACTGAGTTT TGGTTATCAT TGATTAAAGT GTTAACAATT 180  
ATCGTATTCG TTATTTTTTG CTTTTTAATG ATTTTCGGTA TCTTAGGTGG TCATACATAT 240  
25 GGATTTGAAA ACTATACAAA AGGCCAAGCA CCGTTTGTG GTGGTATCTC TGGTTTCTTA 300  
GGCGTATTAT TAGTCGCCCG ATTTTCGGTT GGTGGTACAG AAGTAGTAGC AGTAACTGCT 360  
30 GGTGAATCAG ATGACCCTAA AAAGTCTATG CCTAAGGCAA TTAAACAAGT ATTTTGGCGT 420  
ATTCTTTTAT TCTATGTCTT ATCAATTGCA GTAATTGGTG CAATTATTCC GTACACAGAT 480  
CCATCATTAT TAAGAGCAAG TAGTTCAATA AGTCAAAGCC CATTACAAT TGTATTGAT 540  
35 AGAGtAGgCA TAGCCTTTGC AGCATCAGTA ATCAACGCGG TTATTTTAAC TTCATTATTA 600  
TCCGCTGCAA ATTCAGGTGT TTATACAACA GGCAGAATGT TGTATTCCTT AAGT 654

## (2) INFORMATION FOR SEQ ID NO: 614:

40

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 642 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
45 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 614:

TCGATCTTTA TTGATAATAA TTAAATTGTC GCCTTTAAAA TGTGATATTA ATCCTGCGGC 60  
50 AGGnTGTAACA ACGAGTGATG AACCTAGTAC AACAAGGGTG TCAGCATGTT CAATTTTATT 120  
TAATGCCCTT ATGATGGTAG GTTGATCTAA CATTTCACCG TATAATACGA TGTCCGGTCG 180  
55

ATAAGATTTA TGACATACAT TACAATAAAA ACGATTTAAC GTGCCATGTA ATTCATCAAC 300  
 ATGTTGACTT CCAGCGTCTG AGTGCAAACC ATCGATATTT TCGTGATGA CACCTAAAGA 360  
 5 TTGTTGATTA CGTCTAATT TTGCAATCCA ATCATGAACG ATATTGGGCA TCGTATCGAC 420  
 AAATAGTAAG CGCTTATGGC AGAAATTGAT AAAACCTTCA GGATCATCTT CTAAATAATC 480  
 10 ACGGCTTAAC AAGTATTCTG GCGAAAGCCC ATCTTTTGaA ATTTTCATCA ATAAGCCACC 540  
 CATTGaACGG AAATCTGGAA CGCCACTTGC GACAGATACA CCAGCACCTG TAAAAAATGT 600  
 AATACGATTC GAACTATCTA TAATATGTTT TAGTGTCTCT AA 642

15 (2) INFORMATION FOR SEQ ID NO: 615:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1210 base pairs

(B) TYPE: nucleic acid

20 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 615:

25 TTTCCCCnCC CnCCAAATA TCCAAnGGAA CTTTAATAGT CCAATTGGCA CAGTAAAACT 60  
 ATGGCATTIT ATAAGTATAA TATATCTGTA ATTTATGGTC AATTAGTAAA TTGTTTTTTTA 120  
 30 TTTGAAACAT ATTTACATc AAAATCACAA AGACTTTTAG ATTTTGtTCT AAAAACTCTT 180  
 TAATAATTTA TTTAATGAGA AGAGTTGCTT ATATAGTAAA TTGTGAAGCC GTTAAAACAA 240  
 CGTTACAAAA CCTATATCTT TAATACGGAA CCATATGGTA TGAATCAAGG AATACTTAAA 300  
 35 CTAAAACTTC TCTATCAGAT TTATTTGTTG CGAAATCAAC AACTTTAATT GCTTGCCCTT 360  
 CATTTAATGG ATAATTTGCT TCGGTAATTT TAACTTTTAC AATTTGACCT ATGAGTGATT 420  
 CGTCACCTTC AAATTGTACT TTCATATAAT TATCTGCATA TCCAATAAT GTACCTTCTG 480  
 40 TGtCACCCCTG TTCCTCAGGA ATTACTTCAA GCACATCTTG ATCAAATTTA GACGCATATA 540  
 ACTTTCCGAG TTGATTGCTT AGCGTAATTA ACTTATGCAC CCGTTCATTT TTAATTTCTT 600  
 45 CATCAATTTG GTCATCCATT CTGTCAGCTG GCGTGCCAAT TCTAGGAGAA TAAGGGAAAA 660  
 CATGCAGTTC AGAGAACTTA TGCTTTACGA TAAATCATA TGTTTCTTGG AACTCAGCTT 720  
 CAGTTTCACC TGGGAAACCA ACAATTACAT CACTCGTAAY TGCCAAGTCT GGTAAAGCTT 780  
 50 TATGCAATTT TGTTAATCGT TCTGAAAATC TATCCATTGT ATACTTACGT CTCATACGTT 840  
 TTAATACTGT ATCTGAACCA GATTGTAATG GAATATGCAA ATGACGCACA ACTTTTGTTG 900  
 55 AACGTTCTAA AACGTCAATT ACTTCATCTG TAAGTTGACT TGCTTCAATT GAAGAAATTC 960

CTTTTAAATC TTGACCATAT CCACCTGTAT GAATTCCTCGT CAATACAATT TCCTTATATC 1080  
 CTGAATTCAC TAGTTGCGTC GCTTGTTCAA CTACTTTTTTC CGGATCTCTT GAACGCATTA 1140  
 5 AGCCACGAGC CCATGGAATA ATAcAGAATG TGCAGAAGTT GTTACAACCT TCTTGAATTT 1200  
 TTAATGACGC 1210

10 (2) INFORMATION FOR SEQ ID NO: 616:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 652 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 616:  
 20 TAATAAAATA CAATACTTTT CAATACAGAG AATCCCGCAA TGTCGGGATT CTTATTTATG 60  
 CTGATTTTGT TTTTGTCTAT GATCAGGGAC TTtCAGGGAC TCAATTAATT ATCACAATCA 120  
 TATTTTCTTA TGACCAAATT GATACATATG ATGCTATTTA ATGAGAAAAT TAGGCATCAC 180  
 25 TTGGTTATTG AATTTCTTTC ATTAAACTTT CCAGCTCAAT TTAATAGTTA GTCGACTATT 240  
 ATTCATTAAA CACTTTTTTAA TCATAAAAAA GTGTTTTTga TAATTCACTa CCaAAAACAC 300  
 30 CTTCTTACTT ATAATTCTAT TTGTTACCA TTTCTAATCT TATCGGCTAA ATCATTCACT 360  
 TTTCTTAATC GGTGATTTAC ACCTGATTTT GAAATTGGAC CAGTTGATAC CATTTCTCCA 420  
 AGCTCTTTCA ACGAAATTTT TTGATGTTCT ACTCGAATTC TAGCAATCTC TCTCAACCTG 480  
 35 TCTGGTAAAT TTTCAATACC AATTTCTTTA TCAATCAATT TAATGCTCTC AACTTGTTC 540  
 ATCGCAGCAC TAACTGTTTT ATTTAGATTG GcCGTTTCAC AATTAACGAG TCGGTTAACA 600  
 GAATTACGCA TATCTCTTAC AATACGTACG TCTTCAAATT TTAATAACGC TT 652

40 (2) INFORMATION FOR SEQ ID NO: 617:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 798 base pairs  
 45 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 617:

TnACCACTTT AAAATAGCGC TTAATAAAAtG AAGGGGGCAA GTCTTATGAC GtTTTACaAT 60  
 TTCATCaTGG GTTTTCAAAA TGrTAACACA CCATTTGGTA TATTGGCCGa ACACGTTAGT 120  
 55



TCTAATTACA CAGATCATCA ATTAATTGAA ACTACAAATA GAGCTATTAG CTTATATATG 240  
 GCAAATTAAT TTGAGTAGTA CCAATTATGA TGTATTAGTG CATCCCAAAT ATCTTTTGTT 300  
 5 TTAAGTTTA TTTCATCATT TCTTATCGAA AATGGTGTA TAATGTCTTT ATCTAACCAA 360  
 GTGTTGATAA GTTCATTTGG TACACCATCT AACACATTT CACTTTTACT AATTATAAAA 420  
 CATTCCCAGT CAAGTGAAAC ATTTTGTTGA TTCACATAAT TACATTGATT ATGATTATCC 480  
 10 ATAAACACTC ACTCCTTTAA AATTCTGTAC TCTTCATTGC GTTTTACCCC GTCACATTAT 540  
 CTTTTAAACT AAAATCATCA TTACTTATGA AAAAAATGTA CATCAAAAGC AAAGGTTTTC 600  
 15 GCTACCGAAA AAGTTTAAAT AATGGTtAA TATATtGGT ACTCATTTTA ATAAAAAGAG 660  
 AATACATTTT GAGCTATCAA TACTTTTTAT TGAAGAGGTG TTATTsyTGG CTAAAACGTT 720  
 ATATTTAATG CGCCACGGAC AAACTTTGTT TAATTTTAAG GGAATAATTc AGGGATTG 780  
 20 AGATTGCGCG CTAAACAGA 798

## (2) INFORMATION FOR SEQ ID NO: 618:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 1786 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 618:

ATACGCTCAA TTGATAAAAA TTTAATATCT GCCATTTGTA TACGCATCGC TTCGAATGTT 60  
 35 TCCGTTGCAA TATCAAATGA TAATTGTGAC TGGAACCTTA AACATCGAAT CATACGTAAA 120  
 GCATCTTCTT GGAATCGTTC CTCAGCTATA CCTACAGTTC TTATTATTCG ATTATTAATA 180  
 TCTTGTGAC CATCAAAATA ATCATACAAT TTGTATGCTG TATCCATTGC TATCGCATTG 240  
 40 ATCGTGAAAT CTCGTCGTTG CAAATCTTCG TATAAATCAC GAACAAATGT AACACCACTT 300  
 GGTCTACGGT GATCGACATA ATCTTCTTCA GCCCGGAATG TTGTCACTTC ATAATTTtCA 360  
 45 TCATtAAAAa CTACATTtAT CGTGCCAtGT yCTTnACCTA CAGGTATCGT ATGACTAAAG 420  
 ATAGATTCTA TTTCATCCGG CGTTGCACTT GTTGTGATAT CTATATCATG AATATTTCTT 480  
 CCCATGACAT AATCTCTTAC AGAGCCACCT ACATAATATG CTTCAAAACC ATTGTCTTGA 540  
 50 ATTTGTTCTA ATATAGGCCT TGCCTGTTCA AATAATGATT TATCCATATT ATTACTCGCC 600  
 TTTACTTTTG TTATGCTCAT TTAGCATTTT TTGATAATAA TACTCATATT GATCTGTAAT 660  
 AAGTTCTGAT CCAAAACGTT CAGCAATATC TGCTAGCATG TTTTCTGAA GTTTGTTGTA 720  
 55

	ATCTACGACA AATCCAGTTT CACCATGTTT AATAACCTCT TTAATCCAC CGGCATTTGA	840
	ACCAATTGGA ACGACGCCTG TwTTCATAGC CTCAAGTAAA GTTAGTCCAA AGCTTTCTTT	900
5	TTCACTTAAT AATAATACTA AGTCAGATAA TTGGTAAAAT TCACTTACGC AATCTTGTTT	960
	CCCTAAAAAT AAAACATCCT CTTCTACGTT TAACTCTTTC GTCAATTGAC GCATTGGCAC	1020
10	TAATTCAGGA CCATCTCCAA GTAAAATTAA TTTACTAGGT ATCTTTTCAC GTACTTTTGC	1080
	AAATGTTTCT ATAATAGTAT CTATGCGTTT TACTTGCTA AAATTCGATA CATGTATTAA	1140
	CACTTTTTCA TCTGGTGCTA TACCAAATTG TGATTTTAAT GCTGTGTTAT GTTTAGTTGG	1200
15	AAACTCATTT TCACGTACAA AATTATAAAT CGGTATAATT TCTTTGTTAG TTTCAATAAT	1260
	TTCATGTGTT TCTTGCTA AAGATTTACT CACACTTGTC ACAATATCAC TTTTTCAT	1320
	GCCAAATTTA ATTGCACCTT GGAGTGAATG ATCATAGCCC AAAACAGTAA TATCAGTACC	1380
20	GTGTAGCGTT GTCATAATTT TTATATCTTT ACCTGACATC TCACGAGCTA AAATCCACA	1440
	AATTGCATGA GGTACAGCAT AGTGCATATG CAACAAATCA AGATCATATT CTTTAATAAC	1500
25	TTCAGCGATT TTAGTACTTA ACGTAATATC ATACGGTGGA TACTGAAATA CTGCATATTG	1560
	ATTCACCTCA ACTTGATGAA AAATCATATT CGGTAATGGT TTTCTTATTC TAAACGGGAT	1620
	ATTTGAAGTG ATAAAATGTA cTTTCGTGACC TCGCTCTGCT AATTTAATTC CTAATTCTGT	1680
30	GGCAATAATT CCAGAACCAC CCATGGACGG GTAACATGTT ATACCTATCn TCATTTCGCTT	1740
	GCCCATCCTT TCTTTCTATT TCTCnCTATG ATnCTCGATG CGTAGA	1786

(2) INFORMATION FOR SEQ ID NO: 619:

35	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 844 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
40	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 619:

45	ACAGGTATGG ACTTnGCTCA AATGACACGA CATTATTTAT CAAGACCTAT TGCTATAATC	60
	TTTTGGATCA TTGCAGAACT AGCAATTATC GCTACAGATA TTGCTGAGGT TATTGGTAGT	120
	GCTATTGCTC TTAATCTCCT ATTTAACATA CCTTTAATCG TCGGTGCACT AATAACTGTA	180
50	CTTGATGTAT TTTTACTACT TTTTATAATG AAATATGGTT TTAGAAAAAT TGAAGCTATT	240
	GTTGGTACAT TAATTTTCAC AGTGTTATTC ATCTTTATAT TTGAAGTCTA TATTTTCATCA	300
55	CCACAGTTGA ATGCTGTGTT AAATGGATTT ATACCACATA GTGAAATCAT TACAAATAAC	360

TTACATTCAT CAATTGTACA ATCTAGAACA TACTCAAGAC ATAACAATGA AGAAAAAGCG 480  
 CAAGGaTTAA ATTTGCTACG ATAGATTGCA ACATTCAAGT ATCAATCGCA TTTGTAGTCA 540  
 5 ATTGCTTATT ATTAGTGtTA GGAGCATCAC TATTTTkCAA CTCAAATGCT GACGATTTAG 600  
 GTGGTTTCTA TGATTTATAT CACGCCTTAA AAAGTGAACC TGACTAGGT GCAACAATGG 660  
 GTGCAATCAT GAGTACATTA TTTGCAGTTG CATTATTAGC GTCAGGTCAA AATTCAACGA 720  
 10 TTAAGGTTAC TTTAGCAGGA CAAATTGTAA TGGAAGGATT TTAAAGATTA CACATACCAA 780  
 ATTGGTTAAG ACGTTTAATT ACACGTTCTC TTGCTGTCAT TCCTGTTATC GTATGCTTAA 840  
 15 TCAT 844

## (2) INFORMATION FOR SEQ ID NO: 620:

(i) SEQUENCE CHARACTERISTICS:  
 20 (A) LENGTH: 574 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 . (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 620:

TGTATTGCAG TCATGCCATA AATTTTTTGA ATAATAAACG GAGACGCAGA AATATAAGTA 60  
 30 AATAATATTA CAAATGTCAT ACCTTGAAT GAGCATTGGT AATACAAAAC GTGGCGTCTT 120  
 CAnTAATATT TTGAAGTTTT TAAACATTGT CTTTAAaTCCA CTACTTGACT CACGATTTGT 180  
 CACTGTTAAT GATTCAGGTA CTTTAAATAA AGAACCTATG ACCATGACGA AGCCAAAGAT 240  
 35 AGTCAGAATG ACAAAGACCA TACGCCAGAC AGAATAATTT AAAATTATGC CCCCTATTGT 300  
 TGGCGCAACA ACTGGTGCAA TACCATTAAC AAGCATCAAT AATGCCATAA ACTTAGTTAG 360  
 TTCATTACCA CTATACATAT CACTTGCTAT AGCTCTTGAA ATAAGTGCTG tGCGCCACCT 420  
 40 GTCACTCCTT GAAGaAATCT TAATGCAACC ATCAGCCAAA TATTATGTAC AAAACAATA 480  
 CCTAAACTTG CTAATGTAAA AATAATCATA GCTATAATAA GCGGCTTTCT GCGCCAGTT 540  
 45 GAATCTGAAA TTGGACnAGC AACCAAATTA CCAA 574

## (2) INFORMATION FOR SEQ ID NO: 621:

(i) SEQUENCE CHARACTERISTICS:  
 50 (A) LENGTH: 523 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55

AAGCAAGAGA GAGTACAACA TTTATATGAT ATTAAAGACT TACATCGATA CTACTCATCA 60  
 5 GAAAGTTTTG AATTCAGTAA TATTAGTGGT AAGGTTGAAA ATTATAACGG TTCTAACGTT 120  
 GTACGCTTTA ACCAAGAAAA TCAAATCAC CAATTATTCT TATTAGGTAA AGATAAAGAG 180  
 AAATATAAAG AAGGCATTGA AGGCAAAGAT GTCTTTGTGG TAAAAGAATT AATTGATCCA 240  
 10 AACGGTAGAT TATCTACTGT TGGTGGTGTG ACTAAGAAAA ATAACAAATC TTCTGAAACT 300  
 AATACACATT TATTTGTTAA TAAAGTGTAT GCGGAAATT TAGATGCATC AATTGACTCA 360  
 TTTTCAATTA ATAAAGAAGA AGTTTCACTG AAAGAACTTG aTTTCAAAT TAGACAACAT 420  
 15 TTAGTTAAAA ATTATGGTTT ATATAAAGGT ACGACTAAAT ACGGTAAGaT CACTATCAAT 480  
 TTGAAAGATG GAGAAAAGCA AGAAATTGAT TTAGGTGATA AAT 523

(2) INFORMATION FOR SEQ ID NO: 622:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1871 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 622:

30 AGAAGTGTGA nAAAAATTTA AnAGAGATAT GCACATAGAT GACGCATTGC TATATCCAAG 60  
 CAATTGAGAA AGCTGCTGAT GCTCCAAATC aCGGAATGAG GGAACCATGG AGAGTTGTGC 120  
 ATGTTCCGAA AGACAGATTA GGAGATATGA GTAAGGATAT TTCTAAATTT GCATTTTCCTA 180  
 35 ATGAATTAGA TAAGCAACAA TGTCATTATG ATGCAGTTAC GAAACTAGGT GGCATGTTAT 240  
 TGCTTATTTT AAAACAGAT CCAAGACAAC GTCAAAATGA TGAAACTAC TTTGCATTTG 300  
 40 tGCATATGCA CAAATCTTA TGTTGTTACT TTATGAAGCG GGAATAGGTA CATGTTGGAA 360  
 aTCGCCATTA TATATCTATG ATCCTAAAGT AAGnAAACAC TTGGTATAAA GnAAGATGAA 420  
 GTTCTTGCTG GATTCTTATA TTTAACGGAT TTAGAAGnAG ATATGCCTAA AGCACCACGT 480  
 45 AAAAATAGAA ACTTAATTAC ATTATATTAA TATGTATAAT TATAGAAACA TTAATAAAAG 540  
 CTGAGTCATG AATTGATGGA CATCTATCGA GTTAGAGATT TAATCTAACT TACTAGAGTC 600  
 GGTACAATAA CAGTCTCAGC TTTTATTGT GCAGTATATA CACATTTTTA TTTTAGTATT 660  
 50 TATTtAAAAG TTtCTGCTaA AAATGATTCA ACTTGTTTCA GTGACTTAGC ATTTGCTGAA 720  
 TGAAGGTGTG CAATTTTATC GCCGTTTTTA AATACTAGCA AGCTAGGGAT ACCCATAACT 780  
 55 TCATTTTCAA CAACTACATC TTCTAATTCA TCACGATTAA CAGTATACCA TTGGTAATCA 840

CCTGCCTCAA ATTAAACAAT TACAGGTGTA TCGCTATTAA TTACAGATTT AAATGATTCA 960  
 TTACTTTTGA TTGATTGCAT TGTAACAACCT CCTCTAGATA GTTTAATAAT TTTTATTATA 1020  
 5 GCTAAATTTA TATCATAATA AAAAATTTTA GCTTCAAAAT GAAAGCCTTT TGCTTTGGAA 1080  
 AATGATATAT TTATTTTAAA TACATAAAGG AGGTTGCAGT CGTATGATTA AATTTTACCA 1140  
 10 ATATAAGAAT TGTACAACCT GTAAAAAGGC AGCAAAGTTT TTAGATGAAT aTGGCGTAAG 1200  
 TTATGAACCA ATTGATATCG TTCAACATAC ACCTACAATA AATGAATTTA AAACAATAAT 1260  
 TGCAAATACA GCGGTAGAAA TTAATAAATT GTTTAATACA CACGGCGCGA AATATCGTGA 1320  
 15 GCTTGATTTG AAAAATAAAT TACAAACTTT ATCAGATGAT GAAAAGTTAG AGTTGTTATC 1380  
 ATCTGATGGT ATGTTAGTAA AGCGTCCTCT AGCAGTAATG GCGGATAAGA TAACATTAGG 1440  
 ATTTAAAGAA GATCAATATA AAGAGACTTG GTTAGCGTAA GTGaAATGTA AGCGTTTACT 1500  
 20 AAATATCTCG ATATTTAGAT TCATTACATG TAAATGAAA TAAGCTATAC AATTGTTAAT 1560  
 TTTTATAAAT ATAGTTGAAT AGCATCTAGC CTTATGGCAT CATTAAATGAT GTAAAGATTA 1620  
 ATTAGGAGGG GATTCTCTTG GCAGTACCAA ATGAAYTGAA ATATTCAAAA GAGCATGAAT 1680  
 25 GGGTTAAAGT TGAAGGTAAT GTAGCAATAA TTGGAATCAC AGAATACGCA CAAAGCGAGT 1740  
 TAGGTGATAT TGTTTTCGTT GAATTACCAG AAACAGATGA TGAAATTAAT GAAGGGGATA 1800  
 30 CGTTTGGTAG CGTAGAATCA GTTAAACTG TATCAGAATT ATATGCACCA ATCTCTGGTA 1860  
 AAGTAGTTnA A 1871

## (2) INFORMATION FOR SEQ ID NO: 623:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 451 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 623:

45 TTGCTGTAGA AGATAAAGAG TCCCATCAAT GGATAGGCTT TATAGGTTTG AATTATATTC 60  
 CAGAAACAAG CGATTATCCA TTAAAGAAT TACCGCTTTA TGAAATAGGT TGGCGCTTGT 120  
 TGCCAGAATT TTGGGGAAAA GGATTAGCAA CTGAAGGCGC AAAGGCAACA TTGAAGTTAG 180  
 50 CAGAAGAACA TCAATATAC GATGTCTATA GTTTTACAGC AGAAGCAAAT AAAGCTTCAC 240  
 AACGTGTAAT GGAAAAAATT GGCATGACAG TGTATGATCA TTTCGAATTA CCCAATCTAA 300  
 55 GTAAGTATCA TTTATTaAAA AGGcAAGTGC GCTATTACAT TAATCTTCcG AAAGTGGAAA 360

ATTAGGGGGT GcTTnGtKtn ATTTTTTTAA n

451

(2) INFORMATION FOR SEQ ID NO: 624:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 665 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 624:

15	CGATAGTTAT CTAAAGCCAT TTTnGATTGT GTTATGAAAT CTAATGATGC GTGATAATTT	60
	AATGCrACAT AACGrTaATA TAAAATATCA ATAGTGAACA TTTGAGCAAA TAATGAAGTT	120
	GTTGCTCCCA TATGCATCTC ATTTTCATCA GTTTTCCCAT AAGTTAAAAC AATATTTGAT	180
20	GCCTGTGCTA CGGGATTATC CCTTGTACTA GTAATTGTAA TTATAGGTAT ATGGTAGTCA	240
	TCAATAACTT TAACCATTGA TTGCATTTCA CTTTGCCTAC CATTGTTAGT AATAAGAATA	300
	aCACTGTCGT TCGAATTGTG AGTTGCTAAT AATGTAGCAA AAATATGTGT TTCTTGAACA	360
25	AGTTGAATAT TAAGACCTAT TCTTGATAAC TTTTGGTATA AGTCGGTAGC AACTACAAAA	420
	GATGCGCCAA AACCATATAT AAAAATCGTC TCAGAACGTT TTAAACAATG ACATATTTGA	480
30	TCAATAGTTT TATCATTTAA TTCGTTATTT GCATGATTAA GTGCGCGTGT AGTACGTGTA	540
	TGGAGTTTAG TTCTTAAAGA TTCTGTGCTT TCGTTATTCA TTAATTCAAC ATTGTAAATT	600
	GATGATGCTT TAGGAACATA TTTAGATATA TTTATTTTCA AGTCGTGAAA ACCGCCATCA	660
35	GTAAT	665

(2) INFORMATION FOR SEQ ID NO: 625:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2549 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 625:

	GATGCCAATT AACCGTGCAT ATAATGTTGA GAAGTTAATC GAAGCAATTC AATATTATCA	60
50	AGAAAAACA AATCGTCGTG TTACTTTTGA ATATGGTCTG TTTGGTGGTG TGAATGACCA	120
	ACTAGAACAT GCAAGAGAAT TAGCACATTT AwTAAAAGGC TTAAACTGCC ATGTTAACCT	180
	AATTCctGTC AACCATGTTC CAGAAAGAAA TTATGTGAAA ACGGCTAAAA ATGATATCTT	240

55

	TTCGGATATT GACGCAGCTT GTGGTCAATT AAGAGCAAAG GAACGACAAG TAGAAACGAG	360
	GTAAAGACAA ATGCTAGAGG CACAATTTTT TACTGATACT GGACAACATA GAGATAAGAA	420
5	TGAAGATGCG GGTGGTATTT TTTATAATCA AACTAATCAA CAACTTTTAG TTCTGTGTGA	480
	TGGTATGGGT GGCCATAAAG CAGGAGAAGT TGCAAGTAAA TTTGTTACAG ATGAGTTGAA	540
	ATC <sub>y</sub> CGTTTT GAAGCGGAAA ATCTTATAGA ACAACATCAA GCTGAAAATT GGTGCGTAA	600
10	TAATATAAAA GATATAAATT TTCAGTTATA TCACTATGCA CAAGAAAATG CAGAATATAA	660
	AGGTATGGGT ACAACATGTG TTTGTGCACT TGTTTTTGAA AAATCAGTTG TGATAGCAAA	720
15	TGTCGGTGAT TCTAGAGCCT ATGTTATTAA TAGTCGACAA ATTGAACAAA TTAGTAGTGA	780
	TCACTCATTT GTTAATCATC TTGTTTTAAC GGGTCAAATT ACGCCGGAAG AAGCATTTAC	840
	ACATCCACAA CGTAATATTA TTACGAAGGT GATGGGCACA GATAAACGTG TGAGTCCAGA	900
20	TTTGTTTATT AAGCGATTAA ATTTTATGA TTATTTATTA TTAAATTCAG ATGGATTAAAC	960
	TGATTATGTT AAAGACAATG AAATTAAGCG TTTGTTAGTA AAAGAAGGTA CAATAGAAGA	1020
	TCATGGTGAT CAATTAATGC AATTGGCATT AGATAACCAT TCGAAAGATA ACGTTACTTT	1080
25	CATACTCGCG GCTATTGAAG GTGATAAAGT ATGATAGGTA AAATAATAAA TGAACGATAT	1140
	AAAATTGTAG ATAAGCTTGG CGGCGGTGGC ATGAGTACCG TTTATCTTGC TGAAGATACG	1200
30	ATACTTAACA TTAAAGTTGC AATTAAGGCG ATTTTATAC CACCTAGAGA AAAAGAAGAA	1260
	ACATTAAAAC GTTTTGAACG AGAAGTACAT AACTCATCAC AGCTATCACA TCAAAATATA	1320
	GTAAGTATGA TCGATGTTGA TGAAGAAGAT GACTGTTACT ACTTAGTAAT GGAATATATC	1380
35	GAAGGTCCGA CTTTGTCTGA GTATATTGAA AGTCATGGGC CATTAAAGTGT TGACACAGCG	1440
	ATTAATTTTA CGAATCAAAT ATTGGATGGC ATTAAACATG CGCATGATAT GCGTATTGTA	1500
	CATAGAGATA TTAAGCCACA AAATATATTA ATTGACAGCA ATAAAACGTT GAAAATATTT	1560
40	GATTTTGGAA TTGCTAAAGC TTTAAGTGAG ACGTCTTTAA CTCAGACTAA TCATGTGTTA	1620
	GGTACTGTGC AGTACTTTTC GCCAGAACAA GCAAAAGGTG AGGCAACGGA TGAATGTACA	1680
45	GATATTTATT CTATAGGTAT kGTGTTATAT GAAaTGCTTG TTGGTGAACC ACCCTTTAAT	1740
	GGAGAAACTG CAGTTAGCAT TGCGATTAAA CATATTCAGG ATTCTGTGCC AAATGTGACA	1800
	ACAGATGTAC GTAAGGATAT TCCGCAATCT TTAAGTAATG TCATTTTACG CGCTACAGAA	1860
50	AAAGACnAAG CGAATCGTTA CAAAACAATT CAAGAAATGA AAGATGATT GAGTAGTGTT	1920
	TTACATGAAA ATCGAGCGAA TGAAGATGTC TATGAACTCG ATAAAATGAA AACGATAGCG	1980
55	GTACCTTTGA AAAAAGAAGA TCTAGCAAAG CATATTAGTG AACATAAGTC GAATCAACCT	2040

AGCCAGAAGG TACGGTGTAC GAACCAAAAC CTA AAAAGAA ATCAACACGA AAGATTGTGC 2160  
 TCTTATCACT AATCTTTTCG TTGTTAATGA TTGCACTTGT TTCTTTTGTG GCAATGGCAA 2220  
 5 TGT TTGGTAA TAAATACGAA GAGaCACCTG ATGTAATCGG GAAATCTGTA AAAGAAGCAG 2280  
 AGCAAATATT CAATAAAAAC AACCTGAAAT TGGGTAAAAT TTCTAGAAGT TATAGTGATA 2340  
 AATATCCTGA AAATGAAATT ATTAAGACAA CTCCTAATAC TGGTGAACGT GTTGAACGTG 2400  
 10 GTGACAGTGT TGATGTTGTT ATATCAAAGG GCCCTGAAAA GGT TAAAATG CCAAATGTCA 2460  
 TTGGTTTACC TAAGGAGGAA GCCTTGCAGA AATTAAAATC GTTAGGTCTT AAAGATGTTA 2520  
 15 CGATTGAAAA AGTATATAAT AATCAAGCG 2549

(2) INFORMATION FOR SEQ ID NO: 626:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2286 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 626:

TGCTTACTTC GCCTTCAATA CGTACTAATT CATGTCCACA ACTTGGACAA TGGGTTGGCA 60  
 TATGATATGT GACAGCATCC TCAGGTCTAC GTTCTGGAAT ACTACGTACA ACTTCAGGTA 120  
 30 TGATGTCACC TGCTTTTTTC ACTACAACAC TATCACCAAT TCGAATATCT CTGTCATGAA 180  
 TTAAATCCTC ATTGTGCAAA GATGCTCTTG aTACAGTTGT ACCAGCTACT TTTACTGGTT 240  
 35 CTAAATAGC AGTAGGTGTG ACTACACCTG TTCGTCCAAT ACTTAATTCA ATATCTAATA 300  
 ATTTAGTTAC TACTTCCTCA GCTGGAAATT TATAAGCAAT GGCCCATCTA GGAGATTTTT 360  
 GTGTGAATCC CATCTCATCC TGTTGATCTA AATCATTAACTTAATAACA ATCCCATCAA 420  
 40 TATCATAAGG TAATGACTCT CTTTGGCTTG TCCATTTTTTC AATATACTCT AAAACACCAT 480  
 CGATATTATT TACACGCGCT CTATTTTTAT TCGTTGTAAC ACCTAATTTA TCTAACTCAT 540  
 CTAATGCTTC ACTTTGCGAA CGCGCATTGA AATCAGTGAA ATCATTGACA CTATATATAA 600  
 45 ATACGCTTAG CTTTCGTTTT GCGGTTAATT TAGAATCTAA CTGTCTTAAT GATCCCGCAG 660  
 CAGCGTTTCT TGGATTnGCA AATAACTGCT CATCATTTTT TTCTTTTTCT TCATTTAATC 720  
 50 GTAAAAATGA ACGTCTCGGC ATATATGCTT CACCACGAAC TTCTACATTT AATGGtTCTT 780  
 TCATTTTCAA AGGTATCGCA TGAATTGTTT TTaAATTTTC GGTAATATCT TCACCTGTTG 840  
 TTCCATCACC ACGTGTTAAA CCTTGaACGA AGTATCCATC AACATATTTT AATGATACTG 900

55



EP 0 786 519 A2

	GTTGGTCGAA TTTTCTCAAA TCATCCTCAT TAAATGCATT CCCTAAACTT AACATTGGCG	1020
	TGTCATGGTT GACTTTATTG AAAGAGGCTT GGGCTTCACC GCCAACTCTA ACTGTTGGAG	1080
5	AATCTACAGT CTTATACTCA GGATGCTCCT CTTCTATTTT AATCAGTTCA TGAAGTAATT	1140
	TGTCATATTG ACTATCTGGT ACAGATGGAT TATCCTCTAC ATAGTATTCA TAACTGTATT	1200
	GATTTAATAA ATCATGTAAC TCGTTCACAC GAGACGATAA ATCAGCCATC CCTTAATCCT	1260
10	CCTTTTTTtC AATTGGTGCA AATTGCGCTA ACAAACGTTT TGGCCCTTGT GATTTAAAGA	1320
	TAATATCTAG TTCGATTGAG CCATTTTTCT CGTTTACATT ACTCACCATG CCTTCTCCCC	1380
	AGGCTTTATG CATCACTTTG TCACCTACAT TCCAATCAGA TGACAATACT TGTTTTTTCG	1440
15	TTGACGTTGT TCGTTGACTA AATCCGCGTT TAGCAAAAGG TTTTGCCTTA GGTTGTATCG	1500
	TTTGTCGTTT GCCACTTGAA TGATTTTCTA ATAGTGATTC TGGAATTTCC TTTAAAAATC	1560
20	TGGATGGCAT ATTTGACTGA GGGCGACCAA ATAACATTCT TGATGTCGCA TGAGTGATAT	1620
	ATAACACCTC TTCAGCCCTT GTAATTGCTA CATAACAAAT ACGACGTTCT TCTTGCAITT	1680
	CATGATCATC TTCACTCTTA ATCGCTCTAA TATGTGGGAA TAAAGATTCT TCCATCCCCA	1740
25	TTATAAAGAC AATTGGAAAT TCAAGGCCCT TAGCCGAGTG CATCGTCATT AGTGTTACGC	1800
	CATTTTCAGT ATCTGCCTCA TCAATATCAG CTACTAACGA TAAATCCGTT AAAAAGTTAA	1860
	TTAATGACTG TTCTTCTAAT GGGGTATTTT CCTCATAGTC TTTTGGTACT GACATAAATT	1920
30	CATCGATGTT TTCTAATCTA CTTGAGATT CTAATGTATT TTCACGTTCA AGCATTCTC	1980
	GATAGCCAGA CTTTGTAAAT ACTTCATCAA CTATTTCATG AATTTCTAAA AATTCTTGTT	2040
35	CTTTTATCAA GCTTTGGATT AACTCGTAAA AATTAAGACA CTCTTGTC ACCTTTTTTG	2100
	aCAATCCGAT AAAATCAGCT TCTCCAAGTG CATCAAACAT ACTGATATTG TTTTGAAGTG	2160
	CATAGTTTTG AACTTTTTCA ACAGATGAAG GACCTACACC TCTTTTTGGA ACATTaATAA	2220
40	TACGTTGCAA ACTAATGTCA TCATTACTAT TGGCAATTAT ACGCAAATAA CTTAATAAAT	2280
	CTTTGA	2286

(2) INFORMATION FOR SEQ ID NO: 627:

- |    |                               |
|----|-------------------------------|
| 45 | (i) SEQUENCE CHARACTERISTICS: |
|    | (A) LENGTH: 400 base pairs    |
|    | (B) TYPE: nucleic acid        |
|    | (C) STRANDEDNESS: double      |
| 50 | (D) TOPOLOGY: linear          |

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 627:

55

TGACATTAAA AGTTGGGTGA AnAAACAAGG CATACTTTAG TTAAGCTTGA TGAAAATAAC 120  
 AATGGAATTA ATGCGATTAT TCAAAAAGAA AAAGCAAAAG ATTTAGATAT AAATTATTCT 180  
 5 GCTAAAGGTA CTACCAATTG TATTATTTAG TGGAGAATTA GACAAGCTGT AGCAGCGTTG 240  
 ATTATTGCCA ATGGTGCTAG AGCTGCTGGA AAAGATGTAA CTACCTCCTT ACTTTTTGGG 300  
 GGCTTATGCC nTTAAAAAAG TGCCACCGTT AATGTAAAA GCAGTTGCCA AAATGTTTGA 360  
 10 TThATGTTGC CCCAAAAGAT TTACGATGCC CCTTCCCAAA 400

## (2) INFORMATION FOR SEQ ID NO: 628:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 453 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 628:

AATAATTGG GCGCTnTTTG CGTCGGGATA TTATACCGCT TCCTTAATTG TTCAACATTG 60  
 25 TAATCACTGT TTTTCAATTG ATATTTTGCA GAGTAAATTG GTACTTCTGG GTTATATGAC 120  
 ACTTCGTCCT CTTTATAGTT TTCCAATTCT TTGAAATTCC CGTATTGTAC AAAGAAGTTA 180  
 AATTCCTCGA TTTCTTTTTT TACTTTTTCG TCATCGATTG GTTTTAATGG AATCATTTTA 240  
 30 TTakTTTCCA TTTTCACAGG ATATCTTTTT GTATGATTGT GTGTCATTCC ATCGCTATCT 300  
 TCAACAACTT CTCTAACAAT ATAATGCCCT TTAGCCGTTT TAGTATTTCT GTTAATTTCT 360  
 AAAACTGCTC CTCTrGATTG CAGATTTTCT CCTTTTAATT GGATTTTCAT TTCAGATCTA 420  
 35 ATTAGcCAAG TACCTTTATC ATCTtTTTTA AAT 453

## (2) INFORMATION FOR SEQ ID NO: 629:

40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1221 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 629:

TCATATGCAT TTGCAAAATA AACGCCAGAA GCAAGGTTTA GAATTGGGCC GTCCGTTTTG 60  
 50 CTCAaTTCAC TTGcATTCAA TAATTATGTC TGATCATGAT CAATTGCTT ATCTAATTCT 120  
 GCaATTTTCT TCATTGCTT ATCTGATTG TTTTCTTTCG CCATCATTTG ATCAGACGA 180

55

GCTAATGGTA CTAATACTTT ATCTCCATAT GTGTCAATAA AGTTATAAAA ATAATCATCT 300  
 GTTTTTGATA CAAATCCAGC ACGCTCTTCA GTTTCACGAT ATAAATCTAA GAAAAGATTG 360  
 5 AACTCATCAC GTTCAAGGAA TCTGACTTTA ACACCATAGT TTATCGCTTT ATTAATATTA 420  
 CGTTTACGTT GACTATCAAA TGTCTTTTTC AATGTTTCGG GTGTTTTACC TTCAAGGTTT 480  
 10 AATACGCCCA TCCATCGTAC TTGGCTCGAT GTATCATACT CAGTTGTAAA GCCATGATGC 540  
 TCGTAACCAT GTGATTTAAA CAAGTTTACT AGGGCATCAT TTTTCTCGCG ACCTTCAAAT 600  
 GGCACGATAT CTTTATCATA TAGATGATAT AACCAATACG GATCTAATTT AACATATAAA 660  
 15 CATTGATGTT GCTGTAAATA TTTATCTAAC TCTTTTAAAT AATAATCAAC TAATCCTAAA 720  
 TCTGAAAAAT CCATTACTGG ACCACGATTC GAATAGTAAA CATAACTTCC CATAGTAGGA 780  
 ATTTTAGAGA AAAGGCTTGC TGCAATTACT TTGTTATTGT CGTCTTTAAT ACCTAATAAA 840  
 20 ACTACTTCAA AGCCATCATT CTCACGGGTA ACTATATTTT CTTTACTTG GAAATAATGA 900  
 CTTTCCAATG ATGGATTTTG TACAAAGTTG TCAAATTCGG TAACAGTTAA CTCTGTAAAT 960  
 TTCATGTTTT GATAATTCCT TCCTAAAAAA TTCTGTCTTT AACTTTTTTA AGTGCGGTAT 1020  
 25 ATGCTGCGTA AACAGGTTTA TTAATTGGTT TAATAAAGTC ACCAACATAT TCmATAATTT 1080  
 CAGCATTGTA ACCTTTTTTG AATTAACTA CACCAGCATC TTCAGCATCT TCTGTAAATT 1140  
 TrCCACTAAC ACCATAGAAA TTATAACGGT CAATGCATGA TTTAATGCAT AATTAATCAT 1200  
 30 TTCCCATTC ACTGCATAAC T 1221

## (2) INFORMATION FOR SEQ ID NO: 630:

- 35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1121 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 630:

TGGCCCAaT AACACAAGTA ATTGCTGCTA ATGGTAACAT CACAAAAAAT GAAATCGTAA 60  
 45 CTACAAATGT TAAACCTTGG AATACACCAA CCATTCTGTA TAATCGTTTA CTATAGTATC 120  
 TATTGTGAAT CCAAGTAATA ATAGCTGAAA TAATAATACC ACCTAGAATA TTCGTATCCA 180  
 50 ATGTGGCAAT ACCTGCAATT GATTTTAAAC CAGGTACATT TTCAACGCCT TTTTCTAAAT 240  
 TAGCGCCAAA CGTATGTGGC CATTGTGTTA AAATGGCATT TATAAATGTA TTAAACATTA 300  
 AGTAACCCAT CAATGCTGCA AGTGCTGCAT GACCTGGTGC TTTTITAGCT AAAGAAAGTG 360

55

TTACTGACCA AAATTTAAAC CAAAACGTAT GTTGATCTGC TAAACTCCCC ATGATTGTAG 480  
 GATTTTAAAT TAATGTCGCA AAGCCAAGCA CTATCCCAA GAAAGCGAAC ATTAATACCG 540  
 5 GTACAATCAT TGCACTACCG AAACGCTTTA TCGCATT CAT TCTCTATTCC CTCCATATCA 600  
 TCTTCTCTAA CAATACATCT AATTAGATTC ATTTATAAAT AGATGTCTTA CTATTTAAAT 660  
 ATAATATATA GTAAACGCTT ACACACCTAC AACGACATTG ACGTATTTTG AAAGTATTTT 720  
 10 GTATAATCAG ATTATCTTTT CATATAGTGA AAATTTTTTC ACGACCTTAT ATATGACATC 780  
 GTTGTATTTG TAATACATTC GTTTTAAACG CATAATCAAA CCTATATCAA TACACAAATA 840  
 TATATAATGA CATACAAGAT TTTAATGTAA TAACGATCTA TTACACATTT ATTTTCAAGG 900  
 15 AGGTTGAATA TGTTTTTAGA TGAACACATT AATCGAAACT TTGATAAACT TAATGATAAT 960  
 GATTTACATA TCGCTCACTT TATCAATACA CATATAGATG AATGTAAAAA TATGAAAATA 1020  
 20 CAAGATTTAG CGCAATTCAC ACATGCCTCG AATGCAACCA TTCACAGATT TACACGCAAA 1080  
 TnAGGTTTTG ACGGTnATAG TGGATTTAAA TCGTACCTTA A 1121

## (2) INFORMATION FOR SEQ ID NO: 631:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4005 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 631:

35 AACCTTCCAT TTTACTTGAT CGATAACATC AGTTCGCTT TACAAATCTC TTCATTAATA 60  
 TCGCTCTTAA AACCATGAAA TTTAACATGT TCCGATAAAT GATAATCTTC TACAAGTTGT 120  
 CGATATTCTG ACAAACCATT TCCATGTCCA TAAATATTCA ATTGAATATT GGGATGTTTT 180  
 40 GTTACTAATT GCTTGATTAC TTCAATTGGA TGTTTAATTT GTTTATTTTC AACGAGGCGA 240  
 GCAATTGATA TGATATGATT TTTCTCCTTT TGATTGATGT CAAATTGATA CTTTAAATTT 300  
 GCCACGTAGC CAACCGGAAT ATTGATAACT GGTATTTTAT TTTCAATATA TTGTGAAATA 360  
 45 TCTTGGCATT GCTTTTCTGT TGATACAACA ATCGCTTTAT AACGTGTTAA ATTATTAAAC 420  
 ACTGTTTTAT AAAAATTTT TATACCATTA CCGGCACCGG ATAAATGTGT ACTATGGAGC 480  
 ACAACAATAA CTGGAATACT TTGATTTAAT CCCGCTATAA CATTTCTTAA TTCATGAGGA 540  
 50 CGATCTAATA TGATTTGATC ATTATTTTGA CATAATTGAT GGAGAAAATA TTGAACTAAT 600  
 TCATCTTCTG TATCAAAAAA TTGTTGATGC TGGTCTTCAT TTAAGATAAC CTTTGTGAGC 660

55

EP 0 786 519 A2

	TAGTAATTTT	CGAGTACAAT	CCGTTGTCCT	TCACCTAAAA	TTCGAGAACA	ACTTAAAAAG	780
	CCTCTTCCAT	CATACAATTC	GCGTTTTACT	TTTCTTCTTT	TATGATCAAA	ATAATTCACA	840
5	TAATTTAATT	GATGATACTG	TTTATCTAAA	AAATGAGCAT	ACATTACAAA	TTGCTCTTCA	900
	TCATATATTC	TGACATCATT	TGAATTTTCC	ACAAATTTCA	ATGTGTACCT	ACATGACTTT	960
	TCCCAATACT	GTATCCAGTT	AACTTGCTTT	GTCTTTTTAT	AATTGATTGC	TTTTTGAAAA	1020
10	TAGTCATACA	TTGTAAATAC	ATCATTTTCA	ATCTGATGTT	GCTTCGCATA	TGTGTATGAA	1080
	TAAGGATTCC	ATTTAACATA	TACACATTTT	GAAGATATGC	CGTGTGTTT	GAACAACTTC	1140
15	AATCTATTTA	TTTGCGCTTT	TTCTACACCT	GTAATTTTAC	TTTCTAAAAT	TGTTCTTAAA	1200
	ATGTAATTCA	TATTATCGCC	TCATATAAGT	TTTATTCCGT	ATCTTTATTG	TTTATTTTAT	1260
	ATGAAAAATA	CATCTATTGC	ATGTGTAATT	ATAAAAAAAC	CAGGCCACAA	GGACCTGGGT	1320
20	CATATTGTAT	TATTTGTTTT	GTTTTTTGCG	ACGACCGAAT	AACAATAATG	AACCTAATGC	1380
	TGCAAATAAT	CCACCAAATA	ACGTTGCGTT	ATTTGAGCCG	TTATTTTCAC	TACCTGTTTC	1440
	TGGTAATGCT	TTTGCTTTAT	TGTGATGGTC	TTTAGTAGTA	CTCATTGGTT	TAACAGGTGT	1500
25	ATGTTTTCCT	GCATCCGAGT	CTGAATCGCT	GTCTGAATCA	CTGTCTGAGT	CTGAGTCGCT	1560
	ATCAGAGTCT	GAGTCGCTGT	CCGAATCTGA	GTCGCTATCT	GAGTCTGAGT	CGCTGTCTGA	1620
30	ATCTGAATCA	CTGTCTGAGT	CTGAGTCGCT	ATCTGAGTCT	GAATCGCTGT	CTGAATCTGA	1680
	GTCGCTATCT	GAGTCTGAAT	CGCTGTCTGA	ATCTGAGTCG	CTATCTGAGT	CCGAATCGCT	1740
	ATCTGAATCT	GAGTCGCTGT	CTGAGTCTGA	GTCGCTATCT	GAGTCTGAAT	CGCTGTCTGA	1800
35	GTCCGAATCG	CTATCTGAAT	CTGAGTCGCT	GTCTGAGTCT	GAATCGCTAT	CTGAATCTGA	1860
	GTCGCTATCT	GAGTCTGAAT	CGCTGTCCGA	ATCTGAGTCG	CTATCTGAAT	CTGAGTCGCT	1920
	GTCTGAATCT	GAATCACTGT	CTGAGTCTGA	GTCGCTGTCT	GAGTCTGAGT	CGCTGTCTGA	1980
40	GTCACATCT	GAGTCTGAAT	CGCTGTCTGA	TGTATCTTCT	TCGAAGTATC	CGTTATCAAG	2040
	TGTGAAATCA	TCATGATCCG	TAATTGTTAC	GTCAACTTCG	CCACCATCTG	CATCTTTATC	2100
	ATCTTCAGTT	GTATTTGTAA	CTGTTTGTGT	TAAGCCAGCA	GGCTTTTCAA	AAATAACTTT	2160
45	GTATTTACCG	CTATCTAAAT	TATCAAAGCA	GTATTTACCA	TTTTCATCTG	TTTATGTTGT	2220
	TCCAATTACT	TCGCCTTTTT	CATTTAATAA	AGTAACTTTA	ACATCTTTGA	TACCTTTTTTC	2280
50	AGTTGAATCT	TGTTTGCCGT	CTTTATTACT	GTCGTACCAA	ACATAATCAC	CTAAACTATA	2340
	TTTTGGTGTT	TTATAGAAAC	CACTGTCTAA	TGTCATGTTA	TCTGCATCTT	TAATGACACC	2400
	TGTTGTTGTT	AAACCATTAG	AATCTTTTTT	AGTATCATTT	CCAGAAGTTA	CTGAAGTTGG	2460
55							

	TTGATATTTA CCATTTTCAT CTGTTGTAAC TGTTTTTAAA ACTTTGTCGT TTTCATCTTT	2580
	TAACGTAAC GTTACACCTG AAATGCCCTT TTCATCTTTA TCTTGAACAC CGTTTTTATT	2640
5	TGTATCTTCC CATACATAGT CACCTAAGTT GTAAGTCGGT TTGTAGAAAC CAGAGTCAAT	2700
	AGTATCGTTA TCTTTATCTT TAATGACACC TGTGTGTGAT GTACCATTG AATCTATACC	2760
	TTCATCAGTT CCTGAACCTA CTTGTGTTGG TGTGTAACCT GATGGTGTGTT CGAATTCAAC	2820
10	TTTATAAGTT CCATTTTCTA ATCCAGTAAA TTGATATTTA CCATCTTTAT CTGTTTTAGT	2880
	TGTTTGTAAG ACTTCACCGT TTTCATTTTT CAATGTAAC GTTACGCCTG AAATACCTTT	2940
	TTCAGTTGAA TCCTGCTTAC CATCTTTATT TGTATCTTCC CATAcataat TACCTAAATT	3000
15	ATATTTTGGT GTTTTGTAGA ATCCACTATC TAATGTCATG TTATCAGCAC CATTAAATAAC	3060
	ACCTGTTGTT GTTAAACCAT TAGAGTCTTT TTCAATGTCG CTACCAGATG TTAGTGTAGT	3120
20	CGGTGTATAG CCTTCTGGTG TAGTAAATTC AACTTTATAA TTACCATTAT CTAAATCAGT	3180
	AAATTTATAT TTGCCATCAG CGTCTGTTGT AACTGTTTTT AaCagTTACC GTTTTCATCT	3240
	TTTAATGTTA CCGTTACGCC AGATATACCT TTTTCATCTT GGTCTTGGAT ACCATTTTTA	3300
25	TTTGTATCTT CCCAGACATA GTCACCTAAG TTGTATTTAG GTTTGTAAAT ACCTAAGTCT	3360
	GCAGATAAGT TATCTTTGCC ATTAAGTGTA ATAACTGAAG ATAAGCCGTT TGAATCTAAT	3420
	TCTTCGTTAT TACCTTGTTT TGAAGGGGTT ACTTCATAAC CTTTGGGTAA GTTTGAAAAT	3480
30	TCTACACGGT AATCTCCATT AGGTAAGTTT GGAATCAAGT ATGACCCATC TTCTTTAGTA	3540
	ACTGCTTCTC CTACTTTTGT ATTTGTATTA TTATCAAATA CAGTTACAGT TACATTGCCA	3600
35	ACGCCTTTTT CTCCTAATTC TTGAACACCG TTTTATTAG TATCTTCCCA TACGTAGTTA	3660
	CCAATTTTAT ATACTTCTTG ACCAGctCCG CCACCTTGGT TATTAGTAAA TCCTAAAGCA	3720
	TTGCCAGTAG AAACGGATTT ATTACCTGTT GAAGATAAAG TAGCCATTG AACAAAGTGT	3780
40	GGGCTTTCGC TATTTGTATA TTGGAATTTT GTATTAACCA TTACAACATA AGCAGAATCT	3840
	GCATTTCCAA AATCAATAAC AGCGCTATTG TtGTCGCCAT ATGTAATTTT CTGCAAGTAT	3900
	TGATTTGTTA CATCTGTAAG CTCTTTAGTA TTCACATCGT ATCCTTTATT TAATGTATAA	3960
45	CCTTTAGGAA CTTGATATAT TTTTATATCT GTTACATCTT TATTT	4005

(2) INFORMATION FOR SEQ ID NO: 632:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1440 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 632:

	TATTAGGTTA CTCTAGTTTC CAAAGCGGGA ATTTTAATGT TATTAACAGC AAGGACAGCA	60
5	AAAGCAATAT CGGCGCATTG ATTGAAAATC CAGGAATATA TCCTTTTATG TCTGGATATG	120
	AAAACCTTGAA GTTATTGAAT GAATCAAAAA AACTCAAGA TATCGATAAA ATTGTCTCAC	180
10	AACTTCATAT GGATGAATAC ATTCATAAAA AAGCTAAAAC GTATTCTCTT GGTATGAAAC	240
	AAAAATTAGG AATTGCTATA GCATTTTTTA ATAAACCTCA ATTCATTATC TTAGATGAAC	300
	CAATGAATGG CTTAGATCCA AAAGCTGTGC GAGATGTACG TGAATTGATT GTCCAAAAAG	360
15	CGCAAGAAGG TGTACTTTC TTAATTTTCA GTCATATTTT AAGTGAATTA GTTAAAATCA	420
	CAAACCTCTAT CCTTATTATT AACAAAGGTA AAATTGTTAC AGAAACATCG GAAGAAGAAC	480
	TTAAACAATT TAAAGATAAT GATTTAGAAA ATGTATTACT AGAAATCATA GAAAGGGAGG	540
20	ACCAAGCATA AAATGGGAAC TTAAATTAAA CAAGAATGTT TCAAATTATT TAAAAAGAAA	600
	TCAACTTTTA TCGCACCTAT TGTCTTTATT CTAATAATGG TTGCTCAAGG TTATATTGCT	660
	ACAAAATACA ATGAAATTTT TACGCCACAG GAATCTTTCA CATCTGCTTA TAATGGTTTT	720
25	TCATGGTTTG CATTTTTATT AATTATTCAA GCAAGTACAA TCATTTCAAT GGAATTTTAT	780
	TACGGTACGA TTAAAAATTT ACTCTATCGT GAATATTCAA GAACAACTAT GATTGTTAGC	840
30	AAAATCATCA CATTATTTAT TATTTCTTTA ATTTATTTTG TTATTACAAT TATTGCTTCA	900
	ATTGTTATTG GGTCTTTATT CTTTAATGAT TTAAATATAT TTGAAAGTAG CGGTAATCAA	960
	TTATCTTTAT TGAATCAATT ATTATTAGTT AGTTTAGGCA CATTGTTGG CGTTTGGTTA	1020
35	GTTTTAAGCT TAACGTTGCT ATTATCATCT GCAACAAATT CAACGGGAGT AGCCATTGCT	1080
	GTAGGTATTG TTTTTTATTT TGCAAGTTCT ATTTTAGCAG TTATTCAAAC GGCACTTTTA	1140
	GAAAAAATAG ACTGGCTAAA GTGGAATCCT ATTAATATGA TGAATATTAT GCTTCAAACA	1200
40	GTTGAAAAAG GCTTTAGTAA GTCGACAAAA TTAGAACTTC ATGAATTGTT TATTGGTAAT	1260
	ATTGCTTATA TTTCTATTTT CTTAATACTT GTAGTATTTA TTTTCAAGAA GAAAAATATT	1320
	TAGTAACTTA AAGTATTAAA TGTCTAAATA CACACATATT CCATCGTAAT TCAAAATCAT	1380
45	TTTCAAATCC CTTCACCCAA ATAATGGTGC GGGGATTTTT TCATCCAAAT TTTGGAAATT	1440

## (2) INFORMATION FOR SEQ ID NO: 633:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1323 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 633:

	GCTGACATAA TTGCATCAAA TTTCACATCC CCATAAAAAT CGCTACCACA TAACCTACGA	60
5	TAATACCTAC AAGAACTGGA ATTAAAGATA GGAATCCTTT AAAAAATCCT TGAACGACTA	120
	TTGTTACAAG CAAGGTTATC ATTGCAACAA TTAAGAACT GATATTGTAA CCTTTCATAT	180
	CTCCAGGATT TTCATACATT GCCATATTGA CTGCAGTAGG CGCTAAGCTT AAACCAATTA	240
10	CCATGATGAC TGGTCCAACA ACAACTGGTG GTAATAATTT CATTAAACCAT GCTGTCCCAC	300
	TTAATTTGAT TAGAATCCCG ATGATGACGT ACATAACACC ACTCATGaT AATGCTACAA	360
15	GCATGTCTCC TAAGCTATGC GTACTTAATC CCGTGATAAT TGGCGTGATA AATGCAAAGC	420
	TAGATCCCAA GTATGCTGGT ATTTGCGCCT TCGTTATTAA GATATAAAGT AATGTACCGA	480
	TTCCCGAAGC TAGTAACGCT GCTGATATTG GTAGTCCTGT TAAGAATGGT ACTAGTACTG	540
20	TTGCGCCAAA CATCGCAAAT AAATGTTGTA AGCTTAAAAA TGCCCATTGC GCTGGTTGTG	600
	GTTTTTCATT TACATCTAGT ACGGGTTTTA CTGTTCTGTT AAACATTTCA TCATTTTGCA	660
	TAATATTCAT TTCCTCCGAT AATAAAAAAA TCTCTTTACA TCAGTATATG TAAAGAGACA	720
25	AAAAGTGTGA CAAGTTGCTA CAAGTCATTT TCGTCCATAG AAATTGACTT ATAGTTGTGC	780
	AACATGAGGG TATTATTAGA TAAACAAGCA TATGAAAAC TATTTATCAT TCAACTCCCC	840
	CACCTTTTTC AGTCTCTCGT ACTGAATTAA AAGGGGtATT ATTTAATTAT AACTGCATTT	900
30	CTTTGATCcA TTtCTTCyAA ATAGACACTT ACCGTTTCCT CTTTAGAAGT AGGTAWATTT	960
	TTACCAACAA AATCTGCTCG AATTGGTAAC TCACGATGTC CTCGATCAAC CAAAGCAGCT	1020
35	AAACCAATTT TAATAGGTCT AGCATTTAGC AAAATAGCAT CAAGTGAAGC ACGAACC GTT	1080
	CGACCAGTAT ACAGCACATC GTCAATAATG ATGACTACTT TATCTGTAAT ATCTGTGTCG	1140
	ATGTCTATTG CGTCTTTTGT CGTAAGTGAT GACATGTGcT CTATATCATC TCTAAAGTAT	1200
40	GTAATATCAA TTGTTCCAGT AGGTATACGT TGTGCTCAA TTTGATGAAT TTtATCTTGT	1260
	ATACGATTCG CTAAATATTC ACCTCTTGTh TTGATACCTA AAAGATTAAA TTATCAGTAC	1320
	CTT	1323

## (2) INFORMATION FOR SEQ ID NO: 634:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 761 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear



ACCCATCTCA TCGTATTTTG AATTTAATAG ACGAGATCGA TGTATATCTG AATTCATCCC 60  
 AACTATGGAT TAATGTTGGT ACATCATtAA cGCATAACCA ACATTTTGAG CAGTTGTTTT 120  
 5 ATAAGTAACG TGATTTTTAT CTAATTGCCC TCTTAATGCG TCCTCTGTAA ATTCAACACT 180  
 ATCAGAACCA TTAGAGGTCG CTTCATATAA GTTATTAGAT GCAATATGTG CTAAATCGCT 240  
 ATTGATTTTC AATGGTTTTA ATCCTTTTAA TTTTCTCATT TCATTGCTTA CTTCATAAAG 300  
 10 AGAAATTAAT TGATTTGGAT TTTGCTCAAC TGGACGCTTA TTATGCTCTT CTGACGTAGA 360  
 ATTAGAATTT AATTGATAAG GTTCAATATC TGCTAACATT TCTTTTGTTA AAAATCGTAC 420  
 ACTTAGCACC TTTTTCGATT GTTGATCAGA ATACACTTGT GCATATATGT CGCCATATTT 480  
 15 AATCaGTGTT TGTGTTTTTA AATCTTCATC TGAAAGTTCA AATTCATATT TTTTACCATC 540  
 AACTTTAAAG GACGGTTCCTG GATTAATACT TGTATGATTA AAAATTCTG CAGAATGTTG 600  
 20 TCCTATTTTT AACGGACTAA CATTGACTTT CTCACCTGTA GCATACACTG AAACGATTTT 660  
 TTCACGTTTA GTTGAAACAA TGTAATAACT GTTTTTGTCT TTAAACACAT AATTTTTGTA 720  
 rCCATCTCTA AAAGGGTAGA CrCGATCTGC TTGTCCAAAT T 761

25 (2) INFORMATION FOR SEQ ID NO: 635:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 827 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 635:

nTAAATATAT TTATATATTA TAGAATAGAA AGACCTGAAG ATTGAATATC TTTGCAAAG 60  
 CCTTTAACTG TATCTACTGA TAATTCGTTA ATATCGCGAC CTAAGTTTGT ATTCACTTTT 120  
 40 TTCACAACAT CTGCTGGGCA TGTAATAATA TCTGCACCAA TTTCATCAGC TTGAATCACA 180  
 TTGAATAATT CGCGGCAACT TGCCATAAT AATTTAACGC CGTCTTTACT ATGCGTAACT 240  
 TThACAGCCT CTkTCATTAA TGGTaATGGA TCTACGCCTG TA tCTGCAAT ACGTCCTGCA 300  
 45 AATACTGAAA CATATGTTGG CACACCTTCA GTTACTGCTT CAGTTATTTT TTTAACTTGT 360  
 TCAATTGTGT AAACAGCCGT AACGTTTAAT CTCACATTGT CAGCTGAAAG TTTTTTAATT 420  
 AAAGGAATCG TTGATTCACC TTTTGTATTT ACAATAGGAA TTTTAACAAA TACATTTTCG 480  
 50 CCATATTGTT TTAAAATTGC TGCTTCTTTk TCCATAGTTT CTAAATCGTC TGCAAATACT 540  
 TCAAATGA rA TTGAAGCATC TGGAATTTCT TTCACAGCTT CTTCAGCAAA AGCTTTGTAA 600

55

TTTTTATAAG CTGCTTTCAT TnCTTCAATA TCTGCACCGk CCGCAAATAC TTCTACATTT 720  
 AGTTTAGCCA TATAAYATAG CCTCCTTGAT TCTTATTAAA ATTTTAACAA CATCTGCATG 780  
 5 kCTTTTCTT ACAACCATTT GTAAAAAATG ATTTTATTT CTTtGTT 827

(2) INFORMATION FOR SEQ ID NO: 636:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1478 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 636:

TTAGTCGGTA TAACCATCGG CTAGGTGGTT TTTGTATTAA AAAAGTGGAT aCCAAAATTT 60  
 20 ATTAATAATT ATTTTAATGT TAGAAAAAAA CTAAATAAAA ACTCGCTAAT GATATCCAAT 120  
 AATATGTATA CAAAACGAGA CATATATTGC ATATGATTAA CGAGATACTG AAAATATTTT 180  
 ATCACCCCTA AAATGATTAT TCATTTTCAG CGGTAATTCG ACCTAAAGTC AAACCTACAA 240  
 25 TAAAACCGAT GATAAATACT ACTAATGAAA CGAACCACAT CACGATATTA GTTGGTAAAC 300  
 CTGGAATAC TGCAAGAGG GAGCCAACAA CAAAACCAAT GATTAATGCA AAAGTCATTA 360  
 GTTTATGATG TGTTAGGAAA TACTGGATAA TTTTGCTTGA AATAATGAAT CCAGCAAGCA 420  
 30 CGCCAAATCC GACTGCAAGT AATATAGGAA GACCTGCAAA GTTAAGTTTA ACAACTTCAG 480  
 ATATTGCTAG CATGACCGTA CCATAGACGC CAAATACTAA TAACATAAAT GACCCTGAAA 540  
 TACCTGGGAG TAACATAGCA CTAGATGCAC ACATACCTGC AATAAAATAT TTAATAATAA 600  
 35 GACTAGTTGA TAGAGTAAGT GTTCTCCAG CATGTTTATC ACCATTATTC ATTAATGTAA 660  
 TAACAATTAA GATAGCGATA CCAGCTATAA CCATCATGTA ATGTYTAGTT GTAAATGACG 720  
 40 TTTTATAGTT AGAAATTTTC AATAAATATG GAACGATACC AATGATTAAT CCACCAAAGA 780  
 AAAACATAGT TGGAATATGG TGTGGCTTA ATAAATAATT AAAAAGATTA CTTAGTGATC 840  
 CCATTGCCAG TAACATTCCA ATTATAATGG GGATTAAAAA TGTAAAACTT GGCCAAAAAC 900  
 45 GTCGTGAGAA TATGCCGCTA ATTGAAGCGA TAAATTGATT GTAAATACCT AACATAATG 960  
 CGATAGTCCC ACCGCTAACA CCAGGTACCA AGTCACTCGT TCCCATAGCA AAACCTTTTA 1020  
 GAATATTAAT CCATTTAAAT TGTTCATGA ATAACCTCTT TCAAACGATT GGAATAAAAT 1080  
 50 CATAAATAGC ATCATACCAT ATTACAAATG TCCTAGTGAA ATGATAACAT ATTTTAAATT 1140  
 CATAAAATCC ATTGAGAAAT TATGTGCACT TATTATCATT TATATTTTAA AAGAGAGCGG 1200

AGGTATAAGT AAGTTATAAT TAACTGAACG CATTATTACA AAGTCTTTTT GACTACAAAT 1320  
 TAAAATTATT ATAACTAGT TAAGAAAAC TTTATTTT CGGAGGGAAT ATAAAATGGC 1380  
 5 ATCAACATTA GAAATyAAAG ACCTACATGT GTCTATTGAG GATAAAGAAA TCTTAAAAGG 1440  
 TGTTAACCTG ACAATTAACA CTGATGAAAT ACATGCGA 1478

(2) INFORMATION FOR SEQ ID NO: 637:

10

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1995 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 637:

20 ATTACAGCAT CTTCTCTAGG TAGATTATTA AAAGATAGAG GTCTAAATGT AACAAATTCAA 60  
 AAATTCGATC CATACTTAAA TGTTGACCCA GGTACAATGA GTCCTTATCA ACATGGTGAA 120  
 GTATTCGTAA nGGATGATGG TGCAGAACT GACCTAGACT TAGGACATTA CGAAAGATTT 180  
 25 ATTGATATTA ATTTAAACAA GTTTTCAAAT GTGACAGCCG GTAAAGTGTA TTCACACGTA 240  
 TTGAAAAAAG AACGTCGTGG TGATTACTTA GCGGGAACAG TTCAAGTTAT TCCGCATATT 300  
 ACAAAATGAAA TTAAAGAACG TTTATTACTT GCAGGGGAAA GTACGAATGC AGACGTTGTT 360  
 30 ATCACTGAAA TTGGCGGTAC AACAGGTGAT ATTGAGTCAT TACCGTTTAT TGAAGCGATT 420  
 CGTCAAATTC GTAGCGATTT AGGTAGAGAA AATGTTATGT ATGTTCACTG TACATTACTG 480  
 CCTTATATTA AAGCTGCTGG AGAAATGAAA ACGAAGCCAA CACAACATAG TGTTAAAGAA 540  
 35 TTACGAGGCT TAGGTATTCA ACCAGACTTA ATCGTTGTAA GAACTGAATA TGAATGACA 600  
 CAAGATTTAA AAGATAAAAT TGCATTATTC TGTGACATTA ATAAAGAAAG TGTATTGAA 660  
 40 TGTCTGTATG CAGACTCTTT ATACGAAAT CCATTACAAT TAAGCCAACA AAATATGGAT 720  
 GATATCGTTA TTAAACGTTT ACAATTAAAC GCGAAATATG AAACACAGCT TGATGAATGG 780  
 AAACAGTTGT TAGATATCGT TAATAATTTA GATGGTAAAA TTACAATTGG TTTAGTAGGT 840  
 45 AAATATGTTA GCTTACAAGA TGCATATTTA TCAGTTGTTG AATCATTGAA ACATGCTGGA 900  
 TATCCTTTTG CCAAAGATAT TGACATTAGA TGGATTGATT CAAGTGAAGT AACAGATGAA 960  
 AATGCAGCCG AATACCTTGC AGATGTCGAC GGTATTTTAG TACCAGGTGG ATTTGGTTTC 1020  
 50 CGTGCAAGTG AAGGTAAAAT TAGTGCAATT AAGTATGCTA GAGAAAACAA TGTACCATTC 1080  
 TTTGGTATTT GTTTAGGAAT GCAACTTGCA ACAGTTGAAT TTTCAAGAAA CGTATTAGGC 1140

55

TTACCAGAAC AAAAAGATAT CGAAGATTTA GGTGGTACAT TACGCTTAGG CTTATATTCA 1260  
 TGTTC AATTA AAGAAGGCAC ATTGGCACAA GATGTTTATG GTAAAGCGGA AATTGAAGAA 1320  
 5 AGACATCGTC ATCGTTATGA ATTTAATAAT GACTATAGAG AACAATTAGA AGCAAATGGT 1380  
 ATGGTGATTT CTGGTACAAG tCCAGATGGA CGTTT TAGTAG AAATGGTAGA GATTCCGACA 1440  
 AATGtTTCTT TATTGCTTGT CAATTCCACC CAGAATTCTT ATCTAGACCA AATCGTCCGC 1500  
 10 ACCCGATTTT TAAATCATTT ATTGAAGCTT CATTAAAATA TCAACAAAAT AAATAAATTT 1560  
 GCTAATAAAA CCGGTACTTT CATTGTTAAA CATTGAAAGT ACCGGTTTnT CGTATAATTT 1620  
 15 TAATATTATG TTAGTGACAA GGTATGAAAT AACAATAGTG ACTTTTATAA TTCTAAGTCT 1680  
 CTTGTCATTT CAATCATTTG TGTATAAATG TCATAGTATA CATAATTCAA TGCCATCGCA 1740  
 TGTGGTyGGA CAATCTTATC GTAATCTTCA GTGTAGACTA TAGGTcTTGG TGTAGATAAA 1800  
 20 TCGATAAAAT GTACGAGATG ATCAGGGAAA TCATCTGTTT TAGGTTTGTT GCTTATTAAG 1860  
 ACCACATCGA TATCTAAGTC GATAAGTTTT TGAATATCTA ATGCAACTTG ATyATTATAA 1920  
 AATGGTGCGA ATAATAATAC ACGATCAGTT GAGTCAATTT CTTTAAwkTC TTTAATAGCG 1980  
 25 TaAGTTTnCG GCTAG 1995

## (2) INFORMATION FOR SEQ ID NO: 638:

(i) SEQUENCE CHARACTERISTICS:  
 30 (A) LENGTH: 1107 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 638:

ATTAGTGATG AAAGTCAGAT TGAAGCTTTA TTAACAGCTG AAAAATATTC AGAAATGATT 60  
 40 GGTGAATAAT CACCGTGTA CTCCTTAATC TAAGATTGAG GAGTTTATTT TTAGTCTGAG 120  
 AAAATAAATG ATATGAAAGA AAAATTATTA GGTACTATTA TTTGGAGTAT TGCTACATTT 180  
 TATTATTCAA GAATGATGGA AATAATGAAT TTAGCTATTT TAAAAATAAA AATTGGGGGA 240  
 45 AGTTAATATG CTAAACATTC AAGACGTTAn CATnCTTTCT AAAAAGGAGC AAAAAGCATA 300  
 TAACCGTTTC GTAGAATCTG TAGAAAACGG TAATTTACCA GTACTACCAT GTATTGAAAT 360  
 50 GGATCTAAAA GAGATGCAAG AAGAAACATT AAACCAGAGT AAGATTGGTG GAATGCCATT 420  
 TTTAAAATCT TTTAAAGATA TACCATTAGA TGAAAATAAT GTACCAATGG TATTGCTAGC 480  
 ACAGATTAAT TTGGATGATC TTCCAGAACA ACAAGAATTA TTTCTGTAA AAGAAGGGAT 540

55

AAACAATATA AACTCAAGGC TTGTTTATAT AAAAGAGCCA ATTACAGATT TATCACTCGA 660  
 AAATATTCAA GCGCATTGGA AGTCATTAGA TGCTGATAAT GAGGATATCC CGTTCAGTGG 720  
 5 AGCATTCTTCT ATAGAATTTA GATTGTCGAA ACAAACTATT ACATGTACTG ATTATAAGTA 780  
 CGATGAGGAC GTGCTTGCAT TGTGGAATAA AGTCAATCCA TCCTTCGCGC TAAAATCAAT 840  
 GTTTGGTGGT TATGATGAAT TGATGGAACC TGTGTGTAmC AywTTTACTG CTAAGgAACC 900  
 10 ATTTAATCAA CTTGGTGGTT ATCCATATTT TGACCAAATA GATCCAAGAA CGAACGATcA 960  
 AGAACTGAAA ATGTATGATA GAGTCTTACT GCAAATTGaT TCTACAAGAG ATGGTAATTC 1020  
 15 TTCGATTATa TGGGGTGaTT TAGGTATTGc CAATATCtTA GTGaAATCTA CTGrACCTTG 1080  
 aGGcTATGaa GTTTTGAATG ATTACCT 1107

## (2) INFORMATION FOR SEQ ID NO: 639:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 904 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 639:

30 ATTCATATTA TTATAAATTA TTTCTACACC ATCCCAATTG AGTTGTTTTT CATAATTTAA 60  
 ATGTAATTCC ACTAACTCCC TACCAATTTT AACAAATCCA TATACATCCT TTAATATCGG 120  
 TATTCGCGGA AAACCTTTAC TCAAATCACT TGAATATTTG TTCACATAAT ATTTATGATG 180  
 35 CAAAATTGCA TATATATAAT ATACTATCTC TTCTGAATTA AGATTTATTT TCTTTTTAAA 240  
 AGAATTAGgA AATATTATCT ACaTGCCTCA AACTATCTTT ACCTTTGtAT GTAGCAAAGC 300  
 CTTkGCCATT ACCAATAAAc TGGAAATTAG GTAATATGTC CGTGATCATA GCCGAGAATT 360  
 40 CTTTATTCAT TCCCTGTCCT TGTATATAAA TCACCTGTCC AGTATTCTCC ATTATATTAT 420  
 AATATCTACT TGGCATTTC ATAAATATTT TGTCGTACAC TATCCATTTT TTTGTAAATG 480  
 GTCTATGCAT AAATTTAACA ATTCTCTCTG GATTAATTGA AATATTTTTT CCTTTAGAAA 540  
 45 ATTTTTGGGT AAGTCCTCGT GTCCAACTaA TaAATGTTTC ATCTTTGTTC ACTAAATTTA 600  
 TACGTTCTCT TGAATCTAAG ATATCAATTA ATCTATCTAT TTCAGAATTA TAGTTATCTA 660  
 CAAGTAATTT TGcATTACT AATGCTTTTT CATTCGAAAA ATTTGTTACC CAATTATCTC 720  
 50 TTGCTGaATT TACTCCATTA AATkGAKCTA AATATATAGA ATTTTCAATA TCCTTGGAAAT 780  
 CATAATTGG TAAATAATTC CCCATAGTTT ATGTCTCGGT GATTAATCCA ATCATTGGGG 840

55

AAAG

904

## (2) INFORMATION FOR SEQ ID NO: 640:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 436 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 640:

15	CGATGTCTTT ACTATTAGAC TTAGCCATTG GTTTCACCTC TCCAAAAATT GTAAATGTGT	60
	ATCATCAATA TGAAAGTTAC ATAAACTGA CATATTTCTT TAAATATCA ACGCCATTGA	120
	TAACTTCCTG TTTTAATTGA TACGCTGTAA CAAAATACTA TAGTTAGTGC TTACATGTAT	180
20	ATGTTAAAGC AAGCAGTGGT AAATGTAAAT TATAATTATT CATTAACTTT GCAATATATT	240
	AAATCTTTTA TTCATAGAAG ATAAATATCA AATCAATCAT AATTATTTGA CAACAAATAG	300
	CTAACGATTG TTTTAATCTA CATTGGCTT ATAGCATTTT AAACCTATAC TCTATTTTGA	360
25	TACAATATAA GTGTAAATC AATCATAAAA AGGATATTCA ATATCTGCAT CCAAGAAAAA	420
	CATTACAATT ACCTTT	436

## (2) INFORMATION FOR SEQ ID NO: 641:

30

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 442 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 641:

40	GTTATTAAAT TCAGAGTGGT AGCAAATTAA AGTTAATCAA GAGTTAAGAT GAATTTAATT	60
	CATGAACACG TCTATTATTT TTATAATTGT AGCAAATAAA GCTTTACATC AAGGAGGTAA	120
	TTAAATATGT TCAAAAAATA TGAATCAAAA AATTCATCG TATTAAAAATC TATTCTATCG	180
45	CTAGGTATCA TCTATGGGGG AACATTTGGA ATATATCCAA AAGCAGACGC GTCAACACAA	240
	AATTCCTCAA GTGTACAAGA TAAACAATTA CAAAAAGTTG AAGAAGTACC AAATAATTCA	300
50	GAAAAAGCTT TGGTTAAAAA ACTTTACGAT AGATACAGCA AGGATACAAT AAATGGAAAA	360
	TCTAATAAAT CTAGGAATTG GGTATTATCA GAGAGACCTT TAAATGAAAA CCAAGTTCGT	420
	ATACATTTAG AAGGAACATA CA	442

55

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2472 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 642:

10 CCAATTTTGG TATGAATTAT ACAGATAATT CnGCGCCCGG AGGATCATTT GCTTATTTAA 60  
 ATCAATTCGG TGTGGATAAA TGGATGAATG AAGGGTATAT GGCATAAGGA GAACATTTTA 120  
 15 ACTACTGCCA ATAACGGAAG ATATATTTAT CAAGCTGGAA CTTCAATTAGC CACACCTAAA 180  
 GTTTCGGGAG CACTAGCTTT AATCATTGAT AAATATCATC TTGAAAAACA TCCAGATAAA 240  
 GCGATTGAAT TGTTATATCA GCATGGGACA TCTAAGAATA ATAAACCATT TAGTAGATAT 300  
 20 GGGCATGGTG AGCTTGATGT GTATAAAGCA TTAAATGTAG CAAATCAAAA AGCAAGTTAA 360  
 TAAATCAAAG GAGTTTTTGA TTATGGCAA ATTAGTTACT GAAAACATTT CGAAGCGGTT 420  
 TAAAAATCAA GATGTATTAA AGCATATTAA TATCACTTTA GAAAATAACG AAGTTTATGG 480  
 25 ATTACTTGGT ATTAATGGAG CCGGTAAAAC GACACTTATG AAAATTATAT GTGGCATACT 540  
 TCAACAAGAT TCAGGGGAAA TTAAATTAGA TAATAGACCA ATGACACGAA ATGATTTGCA 600  
 CAAAGTTGGT TCGCTTATTG AAACACCTGC GACATATAAT CATTTAAGTG CACAAGATAA 660  
 30 TTTGAAAATT GTGTGTTTAA ATGAAAGCGT TGATTTCaGC GAAATTAATA GTGTTTTAAG 720  
 CTTAGTCAAT TTAAATGTCG ATAAAAAGAA AAAGGTTAAG GACTTTTCTT TAGGTATGAA 780  
 35 ACAAGACTT GGAATTGCAA TGGCaTTAAT TAAAAaGCCA GAAATTTTAG TATTAGACGA 840  
 ACCATCTAAT GGTTTAGACC CATATGGAAT CCAAGAAGTT AGAGAACTTC TAAAATTATT 900  
 AACAGAACAA GGTACTAGTA TTATTATTTT AAGTCACATT TTATCTGAAA TCCAAGTTTT 960  
 40 AGCAGATCAT ATCGGTATTA TTCATGAGGG TGAGCTAAAA TATCAGCAA GAAATAACAA 1020  
 AGATGAAAAC TTAGAAGAGA TATCTTCAA AATAACGAAA GGTGATTACA AATGATACAT 1080  
 TTAAAGATaG AAGGTATCAA ATTTAAAAaT TCTTTCAGTA TGTATGTTTT ATTAATAAGT 1140  
 45 CcGcTGGTAT TTCTTTGTTT TGCTATTTTC ACAGTCTTAT TCGCCAAAAG TAATACGGGA 1200  
 ACAGCGAATA GTGTGTCACC ATATATAACT TTACTATTTA ATATTTGGCC AATTGCTTTC 1260  
 ATCCCATTG TATTATGTAT GGCTTGTAAT TCGTTATTTA AAATTGAAAT GAGAAATAAA 1320  
 50 TCATTTAATT ATTACTTAAG TAATAATTGG TCGATTACAA AAGAAATAAG AGCAAAGATT 1380  
 TTCATTTTAT CAATAGCATT TTTGGTACAT TGCTTTTTAG TATTTATTAT TGCTTATATA 1440

55

TTGATGTATG TAGTATCTCT ACCATTGATA CCGCTCAACT TTTTATTAAC TCGATACTTT 1560  
 GGTGTGTTTCG TATCAATATT AATAAACTTA GTATTATCAG TCATTGTGT CTTGTTTTTA 1620  
 5 ACATTGAAGA GTTTATTTTG GGTGTTGCCG TGGGGGATAA TGCAGAGAAT CCCGCTTATT 1680  
 ACGCTTGGA TACTACCTAA TGGcTTAGTT GTAAACCATA ATTCAAATA CTTTAATGAT 1740  
 10 CTCAATGCCT TATATATTTT GATTATTGTT AGCATCATTA TTTTCGCGAT AGTAACATTT 1800  
 TTAAATAATA AGAAAAGTTG GCGATTAAAA TGATAATTAA CGAATTAAAG TCATGTAAGT 1860  
 TGAAATTTTC TAAGCAAGCG CTCACATTG TACCCATTAT TGTAACCATA TTGTTTATAT 1920  
 15 TATTTATAAA TTGGTATTTA AACGTAAATT TATGGAATGG TCGACAAATm AGTTTGTTTA 1980  
 CAGCGAGTTT TAATGCAATT ACATCGCTAT TAATTTCTAT AAACGTCTAT CAAGTTATCA 2040  
 ATTTTGAAGA AAATATTGGT CACTTTAATC ATATTTTAGG AAAAGCTAAT AGGCTAAATT 2100  
 20 GGTTAAATGC ATCAATGATT TTTACTTATA CTATTACAGC CATATGTATT CTATTAGCAT 2160  
 CAATTAATTT ATTGTGGCAT TCACATGATA TGAAAATAAC ACTTATGTTT ATAGGcGTaT 2220  
 CATTGTTTTT CAATGTaATT ATATTACTGC TACTTTTTAT TTTTAGTATT TTCATTAAAG 2280  
 25 ATGTAATGGC TATTGTTGTC GGAGTTTTAA TGTTTATTTT TAACGTTTAT TTTGGATTAG 2340  
 AAGTGCTTGG AGATCATTCG TGGTTCTATT TACCAATCAC ATATGCTACA CGTTACGTTT 2400  
 30 ATATGTTTTAn CGAAGGGGAG TATACCAGTT ACATTAACAT TGGGCAATCT ATATnATTAT 2460  
 CACGnTGCCG AT 2472

## (2) INFORMATION FOR SEQ ID NO: 643:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 646 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 643:

45 ATAATATAGG AATTATTTTCG ATnACAGATT TTACGAATCG TGCTACGATG CAAAATGAAAn 60  
 ATAAAGATCC ATATGGCGAA AAGTTAGCTT ATGGAATTGC TTTTAATGGC AGTGTGGATA 120  
 TGCAAGGGGA TAAACAAGTC ACAATTCCAA AATATAGTGT AGTTACAATT ACTGGCGAAA 180  
 50 ATAGTAAAAA TTATCGTGTT ACCGCCGATA ATAAGACTTA CTATGTTAGT AAAGATAAAT 240  
 TAGAATATTT TAACCCGGCA GGTTTATATC AAACGCATAG TTTTAAAAAA TTAGCACCAT 300  
 ATATGAAATC AAATTATAGT AATTACTATG CATACTTTAA TAGTCAATTA CATAAAAAGC 360

55



CACAACAACC GATACAATTA CTTTTCATG ATAATAATCA GTTATACGGT TTTGTTTATC 480  
 CAATTGTAGA TAAAAAGAA TTAAAGATA AGTTTAATAT TAACAATAAC ATTTGGATTA 540  
 5 CTAAAGTTGG GAATGGATAT TGTATTGCCA ATTTGAAAGA AGACAAATGG ATTTATATTG 600  
 AATTGTAGGT GTAAAGATGC TAGATAATAT TATTTATAT TTTAAA 646

(2) INFORMATION FOR SEQ ID NO: 644:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 426 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 644:

20 TAATATCGGA ATTTGATAAT GAAGATATCT AATTTTTTAA TATTCGTAGC TTTTATTTTT 60  
 CTACTTATTT TATGTTTATT TTTAATCTTA CAAATGACAA ACCATTAAAA GTAGCATCCC 120  
 AACATCAAAC AAAAAACAA TTCATCAAAT AAAAATCGCT ACAAACCAA GTCATTAAAC 180  
 25 ACGCAATAAT TAAAATTTTC CACTCATTAT AATTCTGAAT TCCAAATGTC GAATCCGAA 240  
 AACCAAATC CAAATCCAA AAACGCAACT CCAAATTA AAGCATTTCC CTACCATTGC 300  
 GGAAATGCTT TTTACATACT GgATTACTCT GTCATTAAATG ATTTTACAAC GGgAAACCAT 360  
 30 GTCGtCATGT ATGACCaaAG TAGCGTCGCT AtCaTaAgGt GGTTCGGATC TTTATTGGAT 420  
 AATnAT 426

(2) INFORMATION FOR SEQ ID NO: 645:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3241 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 40 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 645:

45 nTTATTTACC CACAACATGT TGCACATTA GGTAAATGGG TACCTTATTT ACTTGGTATT 60  
 GTTATGTTAG GTATGGGATT AACAAATACA CCTAATGATT TCAAAATGGT CTTTAAAGCA 120  
 CCTAGAGCAG TAATTATTGG TGTCTGTCTA CAATTCAGTA TTATGCCAC ATTAGCATTT 180  
 50 ATAATTGCAA AGTCTTTTCA TTTACCACCT GATATTGCTG TTGGCGTAAT ATTAGTTGGA 240  
 TGTGTCCGG GTGGGACATC AAGTAATGTA ATGAGTTATT TAGCCAAAGC TAACGTAGCA 300

	ATATATCTAT TTGCAAATGA ATGGTTGGAA GTATCTTTTCG TGAGTATGTT GTGGTCAGTT	420
	GTTCAAGTTG TATTAATTCC AATTGCTTTA GGTATTGTTT TGCAAATTAT TAATCGTAAA	480
5	ATTGCTGAAA AAGCTTCTAC AGCTTTGCCA ATTATATCAG TTGTTGCTAT TTCATTAATT	540
	TTAGCAATAG TTGTAGGTGG CAGTAAGCAC CAAATCTTAA CTACAGGATT ATTAATATTT	600
	TTAGTAGTTA TTTTACATAA CGTATTAGGG TATACGATTG GATATTGGTT AGCTCGTCTT	660
10	TTAAAATTAG ATCGACAAGA TCAAAAAGCA GTCAGTATTG AAGTTGGAAT GCAGAACTCT	720
	GGTTTAGCTG TGTCAATTAGC aGCATTGCAT TTTAATCCAA TTGCAGCAGT ACCAGGCGCA	780
15	GTGTTTAGTT TCATTCATAA TATAACAGGG CCTATTTTAG CAAAGTATTG GTCAAAAAG	840
	TTATAATTGC ACTAATAGAA TGAAGTGGTC ATCGGACTAT GTTAAGCTTT GATAAAGAGA	900
	AAAAATAGAG GAGTAAATAT ATGTATAGAG CAGTTATATT TGATTTCGAT GGAACAATAA	960
20	TAGATACGGA ACAACATTTA TTTAATGTTA TTAATAAACA TTTAGAGATG CATAATGCCG	1020
	ATCCTATAAG CATTGATTTT TATCGTTCTT CTATTGGAGG AGCAGCTACA GATTTGCATG	1080
	ACCATTTAAT TAAAGCGATT GGTTCCGAAA ATAAAGATAA ACTTTATGAA GAACATCATC	1140
25	TTACTAGTAC AACATTACCG ATGATTGATA CGATTAnATC ATTGATGGCA TTTTTAAAGC	1200
	AACGTCACAT TCCTATGGCA ATTGCCACAA GTAGTGTGAA AGCGGAAATA ATGCCCACCT	1260
	TTAAAGCATT AGGTCTAGAC GATTATATAG AGGTAGTTGT TGGTAGAGAm GATGTTGAAC	1320
30	AAGTTAAACC TGACCCTGAA TTATATTTAT CTGCAGTACA ACAATTAAAT TATATGCCGA	1380
	CACAATGTTT GGCTATTGAA GATTCTGTAA ATGGTGCAAC AGCCGCGATT GCAGCTGGAT	1440
35	TAGATGTTAT TGTTAATACG AATAAAATGA CAAGCGCACA GGACTTTTCT AATGTAGATT	1500
	ATGTAGCAAA AGATATTGAT TACGATCAAA TTGTAGCGCG TTTCTTTACG AAATAGGAGG	1560
	CGTATCATGA TGGGTTACAT TATATTGTTT TTTCTAGCTG GTCCAGTAAT TTTAGGCGTT	1620
40	GGAAATTTGG TGATTGGTCC TATATTTAAC AAACAGACAC CATTTGCGGT GCAAGTAAGA	1680
	TCTTTTGTTG kTGGkTCmAT GrTTTACTTA ATACTCGCAA CAATTGGCTA TTTTTACTA	1740
	TTACAAGGTA AACTTTAACG AGAAAACCAC CTTACCTCAT TAAATGGACG ACCATATGTA	1800
45	TGTGAAATGG TAGAACGTTT ATGTTTATGT ATGAGATAGG GTGGTTTAAA TAGTTACATA	1860
	TATTTTAATA ATAACGTCAC GATGATAAGT ACAATTAAGA TAATATCTAT GCCTACCATA	1920
	ATTGTAGCTC TTGTTGCAIT ACTTCCTTGT TCTTTTGCTG ATTTCATAGC ACGGTAGTTT	1980
50	GGCACAAAGC TAATAATTAG TAAGATTAAT ACAATTACAC CAATTAATGC TGTGTGCATG	2040
	ATGAACGACC TCCTTTATTT TTTTCAATCA ATTCCCAAAT AAACGTAGCA ATCACACCGA	2100
55		

	CAATAATTAA TGCAATCGGT AAAGTCGTAC CGAGTTTAAT CTTGCGCTCT GGAGAATTAA	2220
	TAATAGTAAA TACTGTAAGA CAAATGAGTA TGAAAGCAAG TGTTGCAATA ATAGTTCTTC	2280
5	CAACTAAATA TAGGATGTCA GTTTTTCCCA TACCGATATA ATTTATGATG AAAAATGCTA	2340
	CAGCAAAGAG TACCGATATT TTTGTAGCAC GTAGCAGTAT TTGTTTTAAC ATTGATATAC	2400
	TCCTTTTAA TATTATTAAA ATTATATCAT AATTACCAAG AATAGCTGAA GTTGTATGTG	2460
10	ACTCAACGGT ACTTGAGCAA CTTTTTTAAT TTTTGTAGAA AATCACAAAA TAATTGTTTG	2520
	CAAAGTTGCA AAAGCCTGCT ATAGTAGTTC TGTAACGAT TGCATGGTAT GCAAATATTA	2580
	ATGTACCAAA ATCGATAATT TATAGTATAA TTACGGCAAT AAGTTTTTTT ATGGATTAT	2640
15	TTAGTATCAA TCAGAGATGG GGTAAAGAAGT TATGGAGAAC AATGAACTAC AAAGGGGATT	2700
	GAGTGCCCGT CAAATTCAAA TGATTGCACT TGGTGGTACG ATTGGCGTGG GGCTTTTCAT	2760
	GGGTGCGACA AGTACAATTA AATGGACAGG CCCATCAGTT ATCCTTGCAAT ATTTAATTGC	2820
20	GGGTATCTTT TTATTTTAA TCATGAGAGC AATGGGGGAA ATGATTTATT TAAACCCTAC	2880
	AACAGGATCA TTTGCAACAT TTGCAAGTGA TTATATACAT CCTGCAGCAG GTTATATGAC	2940
25	AGCATGGAGT AATATATTCC AATGGATTGT AGTTGGTATG AGTGAGGTCA TCGCAGTAGG	3000
	AGAATATATG AAGTTTGGT TCCCGGAATT GCCAACTTGG ATTCCTGGTG TTATTGCTAT	3060
	TTTATTATTA ATGGCAGCGA ATTTATTCTC GGTAAGCGG TTTGGAGAAT TTGAATTTTG	3120
30	GTTTGCTTTA ATTAAAGTTG TAACAATTAT TTTAATGATT ATTGCTGGTT TTGGTCTTAT	3180
	nnTCTnTGGT TTTGGAAATG GTGGCCATGC GGTAGGTATT TCTAATCTAT GGACAAATGG	3240
	C	3241

35 (2) INFORMATION FOR SEQ ID NO: 646:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1311 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 646:

	AGGCGTCAAC TCAGATGGTT TAATAATTGC CGTATTACCT GCTGCAATAG CACCGATTAA	60
	AGGTtCGAaC ACTAGTtGAA AAGGATAGTT AAATGGTGCA ATGATCAAAA CTGTTCCATA	120
50	AGGTTCTTTT TTGATATAGC TTTTGTGG AAATAAATAT AAAGGTGTGT CTACATTkTT	180
	TGTTTTAGTC CAGTTTTTAA GTTCcTTACG GGCAATTTTG ATACTTyTCA AAGTTATGCC	240

55

AATATCGCTC TCGTATGATT TAATAGCTTT GCTTAACTTC TTTAATTGCT CTTTTCTAAA 360  
 ACTAATATCT TTAGTTTGTT GTGTATTGAA AAAAGCTTTA CTGTCATAAA ATTTTTGCTC 420  
 5 AATGATATTC ATAATGAAAA GAACCTCCTT ATATGATTAT TTTGGAAAAA GCGATTAATT 480  
 GATTTGAATG TTGTGGCCGT TAATTTTAAA TGGTCTTTCG AATTATATAT GTTGAAAGTT 540  
 GAAAATAGAG CGATGAATCG TGTACATAAT AATATTTATA ACTTTAATCA TAACGAAAAA 600  
 10 GGTAGGAAGA AAACAAAAAT TTATACTCAA CATCGCAAAT ATTTAAGAA AATGTAAAGA 660  
 CAAAAGGGGA ATTGTATAGA AATCACTAAT CTGTGGGTTA GGTAGCTAA AGGAATAAAA 720  
 ACTACTATTG AAAAAGGGTT GTAAATTAGT CAAACGTAAA TAAAAACAG TTCATTGAAA 780  
 15 GTGAAATAAA TTCTACTTTA ATGAACTGTT AGTTAAATAC AACATGTCTA TAATTAGACA 840  
 GTAATATAGT ATTATTTTGT TAATGCTTCA GTGATTTGAG GTACGATTTG TTTTTTTCGA 900  
 20 GAAACGACAC CAGATAAGAA GGCCATGTCA TCTTCTAATT GAACATTGAA TGtTCGCCAA 960  
 CTTTATCTTT TTCAGCACCT ACAACTAAAA TTTTGAATC ACTATTAATG ATGTCAGTAA 1020  
 CAACAAGTAC AAATAAGTCA TATTTTCTT GTGCACTTAC AGCTAACATT TCTTTTTCTA 1080  
 25 rATCTTCTTT ACGATTTAAC ACTTCGTCAA GGTCAACAGC ATTAACCTGT GCAATACGAG 1140  
 TCACATAGTC ACCCATAGTA AATGATTAG CATCCATGTT TAATAAGAAT TCAACTGATT 1200  
 TATCAGTTGT TGAAGCACCT GCTTTtAACA TATCTAAGCC GTACTTTTGA ATATCAACTT 1260  
 30 TAGCAATATC TTTnAATTCT tCAGCTGCTT TAACATCTTG TTGTGTACAT G 1311

(2) INFORMATION FOR SEQ ID NO: 647:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 1498 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 647:

CATATACTTT TTCATTTCTT TACGAGATAC TTTACCAGAG GATTTAGACT TCATACGCTG 60  
 45 ATCCATATGT GCTTGCGTTT CAGaATGTCC ACAAACACAA CGATATACCG CTTCTTTCCC 120  
 TTTACCAAAC AACGTTAATT TCTTTTTTACA GTTTGGACAT CTTGCATTTG TTTTGCGCTG 180  
 TACATTCTTT TTCGTCTTAC AAGATGGATC TTGGCACACA AGCATCTGAC CATTTTTAGT 240  
 50 TTAACTTTA ATCATGAATT TACCACACGT TGGGCATTCT GTGGTTGTTA AATTATCGTG 300  
 TTTATATTTA CGATCACTAT TTTTAATCCC ATTTACAACA TCTTCGTAA AATCTTTCAT 360

55

CCATTGTGcA GTTAAAAGTG GCGACGTTAA TTCTTCTGGT GCTAATTCTA ATATTTGTTT 480  
 ACCTTTTGAC GTTACTTTAA TTTTACCGTC TCTTGATTCA ATGGCATTCA TATTAAATAA 540  
 5 TTTATCGATA ATGTCGGCCC TTGTTGCAAC TGTGCCGATA CCACCTGTTT GTTTTAAAGT 600  
 TTGCGCATAT TTTTATCCT TCAATTGAAT AAAGTTCTGA GGGTTCTCCA TCGCTTTTAA 660  
 TAACGAACCT TCATTAAAAT ATTCTGGAGG TGTTGTTTCA TGTTCTCTAA TATTTGTTTT 720  
 10 TGAAATCTTC ACTTCATCGC CTTCTGAAAA AGGCTGTTGC ATCTCTGTAA TAGATTCAAC 780  
 TTGTCTAATA GATTTAAAC CTA AACAGT TGTTACATTC TCTTTCAAAA CAAATGTGTG 840  
 15 CCCTGCAACC TCTAAAGTTA CAGTTATCGC GTCATACTCG TCGGAGGCA TTAAAGCTTC 900  
 TAAAAACGC TCGACAATCA TATCGTATAA CTTTAATTCT CTATTACTTA AGTCTGACAT 960  
 GACAGGTCTC ACTTCTGTAG GAATAATTGC ATGGTGATCA GATACTTTTT GATTATTAAA 1020  
 20 TATCGACATT TTTGATGAAA ATGTTTTAGA CATTAAATGGG CGTGCTTGGT CTTTATATGT 1080  
 TGTTGCCATC GTCACCTGAA TACGTTCTTT CATAGTATCT ACCATATCAG TTGTTAAATA 1140  
 GTTTGAATCT GTTCTTGAT AGGTTACGAC TTTATGTCTC TCATATAAGC TTTGAAGTGT 1200  
 25 ATTCAATGTT TCTTTAGGTC CAATTTTATA ACGTCTATAC ATATCTTGTT GTAAATCTGT 1260  
 TAAATTGTAC AGTGATTGCG GATACGACTT CTTATGTTTA GTAGCAACAG ATTTAATCTT 1320  
 30 ACCATCGACA TTTTCAAAT TATTAACCAT CTGTTCTAAA GTTCTTTAw TGGcATATCG 1380  
 CTGATTTGaw TCTAGCTGAA AATCAAACCC TTTTACCGTT AATGATAATG TAAAGTATTG 1440  
 TTGTGGnTTG AACTGATTAA TCTCTGTTG GTCGTGTAAT TTAATAAATT GAAACGGn 1498

35 (2) INFORMATION FOR SEQ ID NO: 648:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1044 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 648:

TAAACAATT CGTTTTATTA ATACAGTTTG TAAAAGTATT CTCTTAGTAC TAATAGCTAA 60  
 TGTAATGATA GAAAATACTA GTGTTATTAA AGATTGAAT AAAATAAAAG AAAGTGAAG 120  
 50 ATATTGGAAT GTATTAGATG ATTATTACAC GATTGAATTT GCACCTTATC ACGAAACAAA 180  
 ACAAAGTTTG ATTGATAATA TGGTGCGATC AGAACAAATTA GTAAAGGCTA GTGAAGCAGA 240  
 AAATAATGCG ATTTTATTCA AACCAAAGGG TGAAGTCCGTT GACAATGACA ACTTTTCGCC 300

55

TCAACCTGAT ATTCCGATAA AAAATCAAAA AAATAATGTC GAAGTAATTA TTCCACAAAA 420  
 GTTTCATGCA ATGCGTAATG AAATCAATCA AGCATATCAT TCATGGTTTG AATTTGTACA 480  
 5 AAATAAAAAT AATAAGAGA ATAAGTTATC TATACAGTTT ATCAACAAAA ATGATTGTCTG 540  
 AATTTTTTCA TTTGATGCAC GAGATAGTCG CCATTTGTCA TTTATAGAGG CGCCAATCAT 600  
 TGTGAATGTT CAGGCATCAG ATTTATCGaA TGATTTtAT TATGCCATGa TCaGTCaAGG 660  
 10 CGGGTATTTa TTCaAAAATT ATGaCGCGCT AGTAAAAAAT ATTGGAAAAG TATCCATCTT 720  
 GATGGGGAAA TCCAGTGGAA TAACCAATTA TAAAGATAGC GTGATGGAAA TGTATCATGA 780  
 AAACAATTTG AAATTAACAG TACTCAACTT TTCACAAATC ATTATCGCAA TCATTTTAAT 840  
 15 AATTATTATT TTATTTGATG TGAAATATTA TTTTGAACAG CATCGAAAAT TACTCGTAAT 900  
 CAAAAGCTA TATGGTTATT CAACATTAAG AGCCAATTAC CAATACTTAT TAATAAATAA 960  
 20 TATAGTTGTT ATTTTTATTG GAATATTGAC GAATGTAATT TTACATTCTC ACTATATAAT 1020  
 GATGTTATTT GCAACGATTC TTGT 1044

(2) INFORMATION FOR SEQ ID NO: 649:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 399 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 649:

35 GAACATATTG GGTATGCAA GGnGGTCACT CTTCACACTT ATAAACAACA TTTTAATAAT 60  
 GTAAAGTTTA ACCAGCTAAC ACTTTTGTTA GCTGGTTTTT ATTTTCCTTC AATTTTAAAA 120  
 TGGTTAAGTC CCCTTCTATA TCTTATAAGA CAATCATTAT AATCAATTCA ATTAATACAT 180  
 40 TAACAACCAC AACTAATAAA TATAGTAACT TCAAAATCCA TATTTATGTC TAAAGATAAT 240  
 CTCAATGTTG TTCACGTCAA TAAAATTATC CCTAGGTTTT TAAAAATTGT ACATGTTTAA 300  
 ACAATCAAAA GTGTACATTA TTAAATTATC ATTTCCAGTT AGATTTAGAA AACATTCACA 360  
 45 CCACGCATGG ACCAACGTAT TCGTCTTcAT TcATTTTAG 399

(2) INFORMATION FOR SEQ ID NO: 650:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 747 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 55 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 650:

	GTACTAGGTT CTAnGATTCC CTA	CTACTGTAGT CATAGATATT TTCCAACAG AGCCTGACAA	60
5	CTTTTCATTT TTAATAGTGA TCACTATATT AATCGCTATT GCAATTTATG CATGTCGATT		120
	TGTTTGGGTT TATTTCTGGT ACAAAGATTT TTATTTCCCG AAAAATATAC AATCTTATCT		180
	AGACGAGGAA CATGATTCAC ATGAAACACC ACCTTCTCGA GTGCGTTACG CATTTATTAT		240
10	GACCATGTGT GGTATTCACG GTACAATTTT ACTTTCAATG GCACTTACAT TACCATTTAT		300
	CATTACAAAA GGACAAGCAT TCGAATACCG TAATGATTTA TTGTTTATTG CATCTTTCAT		360
	GGTATTAATT AGTTTAATCT TAGCGCAAAT TGTTTACCT TTAATTACAC CATCTGCCGA		420
15	AGATACTACT TTAAAGGTA TGA	CTTATCA ATCTGCCAAA ATTTTCATTG TTCAAAAAGT	480
	GATCCAGCAT TTAAAAACG AAAGTAAAA AGACAAAAAC GATACAAATT ATCGCCCAGT		540
20	ATTAAACCAA TACTATGGAG AATTGTTATT TTTATTAAAT TCAGAACCTG ATAATCAAAA		600
	TACrAAAGAA CTCAAACGTT TAGAAGATAT TGCAAAAGTA ATCGAAACAT CTACACTTGA		660
	GCGTTTAATT GATAAAGGTA AGGCAACATA TCAGGATATT AATAATTACC GCAATATTGT		720
25	CGAATTAACA GAGACACACC GTACTGC		747

(2) INFORMATION FOR SEQ ID NO: 651:

(i) SEQUENCE CHARACTERISTICS:

- |    |                             |
|----|-----------------------------|
| 30 | (A) LENGTH: 1373 base pairs |
|    | (B) TYPE: nucleic acid      |
|    | (C) STRANDEDNESS: double    |
|    | (D) TOPOLOGY: linear        |

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 651:

	ATACAATACT CTTTTATTAT TCAATAAGCC ACTTCCTATA GCAAATGTTT AACTTTAAAT	60
40	ATTTTTCGAT GCTAACAAAA AATCACACTA TCATCTTTTA AAATGAAAGT GTGATTACAA	120
	GCAAATCTGT AAAATTATA AAGCAGAAAC AATTCAACTT TATCATTATG ACATTTCAT	180
	TAAACCTTCT ACATTATAGT TCCAAGCATC TTACACATGA ATGCAAGTAT TTAACGATTT	240
45	AATTGTGACA TAGCCTGTTG ATATTGTGTT TCATTGATAT AATTTTGTTG CTTCAATTTT	300
	TCTAAGTTCTG TGCTTACACG TTGCGTAAAA TTCTCTGACA TATTATTGAT ATTATATACG	360
	CTAGGTGCAT TGACTTTACT AGCTAAAATA GCGCTTTGTA AAAGTGTAT GTGAGACATT	420
50	GTTGTACTAT TTTTATTCAC GGTGTGTTCA AAGTAATGGT TTGCTGCGCC CTCAAGCGTA	480
	TATTGATTAT CCCCAGTA AATATTATTT AAATAAAGC TTAAATTTT GTTCTTATTA	540

55

TCATTATCAT AAAAATAATT TTTGACAACT TGTGTGTAA TGGTACTACC ACCTTGCACA 660  
 TCTCTGTCGC TAATCGTTGA AAATAAAGCT CTAGTTGTAC CTTTCAAATC GAATCCATGA 720  
 5 TGATTGTAGA ATCGTTCATC TTCCATTGAA ATAAAGGCAC CTTTAACATA CTCTGGCATG 780  
 TTATCAGCTG ACACAAAACCT ACTTTTATTT TCAATTTTTC TTAGTTCATC CACATTATCG 840  
 CGTGTAGATA AAAAATACAT GATACCAATA AACAATGCGA TAATGATTAG AATGGTTAAT 900  
 10 AATATTTTITA ATAGTATTCG TTTACTTTTT TTCTTTTTTC GCGGTTTGCC AACTGGTTGA 960  
 TAATACGTAT TATAGTGAGG TTCGTGTTTC ATATGCTCAA AATGTTTCATT TGAGTTTGAG 1020  
 15 TACCTATCGC TTCTTTTCAT GCGTTTGCTC CTTCTTTTAA AACTCACTTA GtATATACCT 1080  
 TGaGTTTACC AGTACTATCA CAAATAGGCT ACACTTTTTG GGAAAATCAG TCCAAGGGCT 1140  
 TACAATCGTA TACGCCATCA TACTTACTTT TTTGTTTTTT GAAAAAATTA TAGATAAATC 1200  
 20 ATTGCAATTT TAAATATTAA TCATGTCAA TATTGTTATA TTTTATAAAA ATAAAAGACC 1260  
 ATCCCTATTA AATGCCAATA GAGACGACCT TTTATTTGTT ATTCATTTAT TAAAACTAAA 1320  
 ATCCATATTT CATTTCAAAC GAAAATATAT AAATTTTAAC AATCGrTAAC CAC 1373

(2) INFORMATION FOR SEQ ID NO: 652:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 859 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 652:

GGCAGATAAT TTAGTCATTG TTGAATCGCC TGCAAAAGCA AnAACCATTG AAAAGTATTT 60  
 AGGTAAGAAA TATAAAGTTA TAGCTTCAAT GGGACACGTC AGAGACTTAC CAAGAAGTCA 120  
 40 AATGGGTGTC GACACTGAAG ATAATTACGA ACCAAAATAT ATAACAATAC GCGGAAAAGG 180  
 TCCTGTTGTA AAAGAATTGA AAAACATGC AAAAAAGCG AAAAACGTCT TTCTCGCAAG 240  
 TGACCCCGAC CGTGAAGGTG AAGCAATTGC TTGGCATTTA TCAAAAATTT TAGAGCTTGA 300  
 45 AGATTCTAAA GAAAATCGCG TTGTTTCAA CGAAATAACT AAAGACGCTG TTAAAGAAAG 360  
 TTTTAAAAAT CCTAGAGAAA TTGAAATGAA CTTAGTCGAT GCACAACAAG CGCGTCGAAT 420  
 50 ATTAGATAGA TTAGTTGGCT ATAACATCTC GCCAGTTCTT TGGAAAAAAG TAAAAAAGG 480  
 GTTGTACGCG GGTGAGITC AATCTGTTaG CmTTCGTTTA GTCATTGACC GTGAAAATGA 540  
 nATTGAAAC TTTAAACCAG AnGAATATTG GACTATTGAA GGAGAATTTA GATACAAAAA 600



aAAAGATGTT GAGAAAATTA CAGCTGCATT AGATGGAGAT CAATTCGAAA TTACAAACGT 720  
 GACTAAAAAA GAAAAAACGC GTAATCCAGC AAACCCATTT ACAACTTCTA CATTACAACA 780  
 5 AGAGGCGGcA CGTAAATTAA AcTTTmAGC AAGAAAAACA ATGATGGTCG CACAACAATT 840  
 ATATGAAGGT ATAGATTTG 859

## (2) INFORMATION FOR SEQ ID NO: 653:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 747 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 653:

20 TTCAACTTCG TTTGGAAATC ATGTTCTCTCA ATTGTTGGAC TTAAACGAAT TTCTTTAACA 60  
 TTGATAATTT TTTGTTTCTT TTTCATTCTT TTTTCTTTT TCTGTTGTTT GAATTTGAAT 120  
 TTACCGTAAT CCATAATTCT TGCAACTGGT GGTTCGCGAT TCGGTGCAAC GACCACTAAG 180  
 25 TCTAAATCTA CACGTTTCAGC CATTTCTAAA GCTTCACGCT TTGATTTAAC ACCAATTGT 240  
 TCACCATCTT GACCGATTAA ACGTAATTCT TTTGCACGAA TTTTGTCATT GATTTGAGTT 300  
 TGATCTTTTG CTATGGTTGA CACCTCCAAA ATTTTACGA AATTGCACC AAGCAAAAAG 360  
 30 GAAGAGCAGG TATAAAATAC CCGCTCTTCC TTATACACAG TTATGTGTAA TGTGATTAAC 420  
 CTGCCAACTG CTTTATGCGT CGCTACAGGT GAGAAGCGGG TGCTTCTACT TGGTTCGTTT 480  
 35 CGTATTCAAC GTTATTAATC ATATCAACAA TTCACATTTA AGTCAACACT ATAACGTAA 540  
 TTATTTTAT TTAAACCTTT TATTTTCATCC ATTGACACGT CTTGACGTAA ATCTACTTGT 600  
 TCTAATGGAA TTTTTCGTT TTTATATCGA AGCTTATGAT AAATAAAGAA TGCTAAAAAT 660  
 40 ACTGGAATTC CCATATACGT AATTAAGAAG CGACTAAAAT TAAAATCTCC TGTnTTAATA 720  
 AAGTCAACAT CTTGCCCAAn AAnTACT 747

## (2) INFORMATION FOR SEQ ID NO: 654:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 501 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 654:

TTGTGCTTCT TTTTtagCTT CTTGAACCTC TTGTGCCTCT TGTGATGTAT CACTyAAATT 120  
 ATTTGCACTT GCTTCTTCTT TTATCGCTGC TTGTTGTGCT TTCAATGCCA CTGCTTTTGr 180  
 5 TTCTTyATTT GATACAGCCA CACTTTTATC CGCTTCTGCT TGTGCTTCTC TTTTAGCTTC 240  
 TTGAATCTCT TGTGCTTCTT GTGATGTATC ACTTAAATTA TTTGCACTTG CTTCTTCTTT 300  
 TATCGCTGCT TGTTGTGCCT TTAATGCCGC TTGCTCATTT TTAGATTTGT TTAAAAATCC 360  
 10 TTCAACACGT TCTTTTGTAT AGGCAACCGT TTCTTCAAGT TCGGTTTTTC TTTCTTCAAA 420  
 CTTTTCGAC AGTTCTTGTh CTTTGACTTT nAAATCATCT GCTTTTTGAT AAACTTTATT 480  
 15 TTAAATACC AACCTAAAGC C 501

(2) INFORMATION FOR SEQ ID NO: 655:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 830 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 655:

CACCTTGTCa TAATTAATTT TTTGATTTTT CATTTTACTG ATAATAGGTT CAGCATTaAT 60  
 CATGATTTAA CCTCCACAT TTAATCATTa ACTTCTATTa TATATGATTC ATATTAAATG 120  
 30 TCAGTCAAAA AAGTTAGAAA TTCATTTTAA TGCATTTATA TTTCGAAAAT CCCTTATGTa 180  
 TCTAAAAGCA TTTTTTAAAC TTGAATTTTa AAACACTAAA CAACACATAC GTCTCTGTGT 240  
 35 CATTTTCATT TTTTGTATGT CATATATATG TTTACTTCAT TTAAATCAAT TTCATCTTAT 300  
 AATTTATCGT GTATTTTACA AAAGATTGAC TTCAATTCAT CGTAAAAGTT ATACTTTTGC 360  
 CATTTTTTAA TGTAACATGG TGTTAGTAAT AAAAATAATA CATTGAGGTG TTTTACATGa 420  
 40 CAGCATTATT CCCTTATATC GCTTTTGAAA ATTCAAAAGA AGCCCTTGCA TATTACGAAG 480  
 AAGTATTTGG TGCAACTGAC GTTAAACGTT TAGAAGTTGG CGAAGAACAa GCGTCACATT 540  
 TTGGTATGAC TAAGGAAGAA GCGCAAGAAG CAACTATGCA TGCTGAATTT GAAGTGCTTG 600  
 45 GCGTAAAAGT GTTATGTTCT GATTCTTTTG GTCGCGCTGa CAAAATTaAT AATGGCATAT 660  
 CATTATTAAT TGATTATGAT GTTAACAATA AGGAAGATGC TGATAAAGTT GAAGCATTCT 720  
 50 ATGAGCAAAT TAAAGATCAT TCTTCAATTG AAATAGAATT ACCGTTTGCT GACCAATTCT 780  
 GGGGTGGCAa AATGGGCGTC TTTACCGATA AATACGGTGT TCGTTGGATG 830

(2) INFORMATION FOR SEQ ID NO: 656:

55

(A) LENGTH: 539 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 656:

10 GTATCCATGG GCCCGTTnCG CACAACATTT GgnACAATTA GGTACACGnG TTGTCaTCGG 60  
 TCGTTTCGGT ATAATTTTAT CGAATGaAGG CGGTGCGTTA CAAACAATGA AACTACCATA 120  
 CGAATATTAC ATTGGTGGTA AATTAGGTTC TGGTCAACAA TGGTATTCAT GGATTCATAT 180  
 15 CAATGATTTA ATTCAAGCTA TTTTATTTTT AATAAATAAC GAGTCAGCTA GTGGTCCGTT 240  
 TAATTTAACT GCACCTATAC CTGAACGTCA AAATTTATTT GGCTACACTT TAGCAAGAGC 300  
 TATGCATAAG CCTCATGAAA CTTGGGCACC AAGTCTTGCA ATGCGTCTCA TACTTGGTCA 360  
 20 AATGTCAACA GTAGTATTGG ATACTCAAAA AGTATTACCT aATAAAATTC AAGCATTGGG 420  
 aTTCCAATTT AAATATAGTA ATTTAAAAAT GgnACTTGAA GATTTAATTA AAGAATAATC 480  
 AATACCATTA ATGAGCATTa GAAACAACAT ATGTACTAAA TGTAATGTCT AGAGCGACT 539  
 25

## (2) INFORMATION FOR SEQ ID NO: 657:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1106 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 657:

ATCAAGTTGA ATATCATCCA TATTTAACCC AACATAAATT GAATTATAT TTGGCAGCAC 60  
 AACGTATCGT gATGGaATCT TGGTCACCAT TGATGAATGC ACAAATTTTA AATGATGAGA 120  
 40 CAATTAAAGA CATTGCTCAA GAATTAGGAA AGTCACCTGC CCAAGTTGTT TTAAGATGGA 180  
 ATGTGCAGCA TGCTGTGGTT ACAATCCCTA AATCGGTGAC ACCAAACAGA ATCTCTGAAA 240  
 ATTTCCAAAT ATTTGATTTT GAATTATCAG ATGAACAAAT GACGCGAATT GATGGTTTAA 300  
 45 ATCAAGATAA GAGAATTGGA CCTGATCCAA AAAAATTTGA AGGCTAGATT AAAATCGCTC 360  
 AACTGATGAA AAGGTTAGAT GAATTGTCAG GGCTTGGGAC ATTAAGTTCT TAGGCAATGT 420  
 50 AAAAAAGCTG ATTTCTATTA ATTATTTGAT AGAAATCAGC TTTTGTGATA TGTATTTTAT 480  
 AATGTACAGC TCGTTGAGCT GCTATTTTCC TTATATTAAG TGCCATTAAT ACAAACCTA 540  
 GCTCTCGTTT AACTTTATTT AtTCCTCGAA CTGACGTTCT AGTTAAACCC AAAATAGCCT 600

55

CTGGTTCAGA AAGCTTTTGA TTAATTTGGA CTTTAAAGTA TTCCCAATTA TAATTCTTCA 720  
 TGATTtTCTT ATTGGATTTT GAATTTGGTT TCATGCATTG TTGCCTCAAA GAACATGCTG 780  
 5 AACAGTCATC GCATTCATAT AGTTTGAAGT CTCGTTTAAA ACCATATCTA TCATTACGGT 840  
 ATGCATATCT TTTAAACCT ATTCTTTTGT TATTAGGACA TATAAATTCA TCATTAAGTT 900  
 CGTCATATTT CCAATTTTGA GTGTCGAAAA TGTCACTTTT AAACTTTCTA GTTTTATCtT 960  
 10 TAATAACAT GCCATACGTA ATAAGTGGCG TTTTATTAAA ATCATCTATA ATGGCCATAT 1020  
 AGTTTGGCT CACTACCCAT AACCTGGCAT CAGCTACCAA ATGAACCGAA GGGATTTTTG 1080  
 15 GAATCCATTG GTTGAAAAAA TGGGAA 1106

(2) INFORMATION FOR SEQ ID NO: 658:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 658:

TTTTAACTTT ACTCTTTGAT TTAAAGAGTG ATAAATGTTT ACAGTTTAAT TAAAACTGCA 60  
 TAAGAACTTC TAGCTTTTCT CTTTCGTTCA AAGAGAAGCA GCTGTTTCGA GTTTAATCAA 120  
 30 AACCACATAA AGCTTTTAAC TTTACTCTTT GATTTAAAGA GTGATAAATG TTTACAGTTT 180  
 AATTAAACT GCATAAGAAC TTCTAGCTTT TCTCTTTCGT TCAAAGAGAA GCAGCTGTTC 240  
 35 GCAGTTTAAT CAAAACCACA TAAAGCTTTT AACTTTACTC TTTGATTTAA AGAGTGACAA 300  
 ATGTTTACAG TTTAATTAAA ACTGCATAAG AACTTCTAGC TTTTCTCTTT CGTTCAAAGA 360  
 GAAGTTCTAA TACCACCATA TCGTGCGATC GGGAACGGTA 400

40 (2) INFORMATION FOR SEQ ID NO: 659:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1899 base pairs  
 (B) TYPE: nucleic acid  
 45 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 659:

ATAATTACAA TGGCTACTGC AATTATTGTT GGAATTATAT ATTTCAAGTT AGTCATCATT 60  
 TATGCTCCCT TGATTTCAAA TTCATATCAT TAGTTTACCA TATTGAAGAT GATATAATAA 120

55

	AAC TT TAACT GAG CTT CATG GGG CACCAGG TTT TGAAGAA GAAG TAAAAA ATTATATGAC	240
	TCAGCAAATG GCGCCGTACG TAGATGAATT TATTGAAAAT CGTATGGGTG GATTTTGTG	300
5	TGTGAAAAAA TCTAAAAATC CAAATGCAAA ACGTGTAAATG ATTGCAGCAC ATATGGaTGA	360
	AATCGGATTT ATGATTACAA ATATCACTAA AAATGGAATG ATTCAATTCA CAAATTTAGG	420
	TGGTGTTCGA AATGATATTT GGCAAGGACA ACGCTTAGTA ATTAAAAATA GAAATGGCGA	480
10	TAAAAATTATC GGTGTGTGTTT CTAATATACC TAAACATTTT CGTACTGGTA GTGAAGGTGC	540
	ACCGGAAATT AAAGATTTAA CATTAGATAT AGGTGCTCAA AATGAAGATG AGGTGCGTGA	600
15	gCGCGGAATA GATATAGGAG ATACAATTGT ACCTCACACG CCATTACACAC AGTTATCTGA	660
	ACATCGATAT AGTGCTAAAG CATGGGATAA TCGTTATGGT TGTGTCTTGG CAATTGAAAT	720
	ACTAGAATTA TTAAAAGATA TAGAATTAGA TGTAGACTTG TATGTTGGCG CAAATGTTCA	780
20	AGAAGAGGTT GGATTACGAG GTGCGAAAGC ATCTGCAGAG ATGATAGACC CAGACGTTGC	840
	ATTTGTAGTT GATTGTTTAC CTGCCAATGA CGTTAAAGGA AACCAACCAT TATCTGGTGA	900
	ACTTGGTAAA GGGACGTTAA TTCGCATAAA AGACGGTACA ATGATTTTAA AGCCTGTATT	960
25	TAGAGACTAT TTATTAAAGT TAGTAGAAGC ACATGACATT GAACATCAAT ACTATATGTC	1020
	ACCAGGTGGA ACAGATGGTG GAGAAATTCA TAAAGCTAAT ATTGGTATTC CGACTGCAGT	1080
	TATTGGTGTA TGTGCACGAT ATATTCATAG TACAGACTCA GTATTTGATA TAAGAGACTA	1140
30	TTTTGCAGCT AGATCTTTAC TTTCAGAAGC CATTTGTAAT TTAGATAATA ATCAAATAGA	1200
	AACATTACAA TATAAATAAT CGGGTAATAA CAACTATTAT CTCTAAATAG TTATATATAA	1260
35	TCATTAATTA AGGAGACATA AAAATGAAAC AACTTGAATC AGAACACAA TTTGAATCTT	1320
	TAAAACAAGG TGCTACAGTA TTTGAATTCA CTGCAGGCTG GTGTCCAGAT TGTAGAGTGA	1380
	TAGAACCAGA TTTACCGGAA TTAGAAGCGA GATATCCTAT GTTTGACTTC GTATCAGTAG	1440
40	ACCGTGATAA ATTTATGGAT ATTTGTATTG AAAATGGTAT TATGGGTATT CCAAGTTTTC	1500
	TAGTATATAA AAATGGAGAA CTGCTTGGA GTTATATTGG AAAAGAACGA AAATCAATTG	1560
	AACAGATAGA TGCATTTTTA GCTCAATACG TGTAATTTAG ACTAGAGAAA AACGGGGTAA	1620
45	TACTCGTTTT TCTCTGTTAC TATGTGTTGA TTTATTGTAA ACTATTATAA GGTGCGAAAT	1680
	TAGGAGTGTT ACATATGAAT ACCTTTCAAA TGAGAGATAA ATTAAAGGAA CGTTTAAGCC	1740
50	ATTTAGACGT TGATTTTAAA TTAAATCGTG AAGAAGAAAC TTTGCGTATT TATCGAACAG	1800
	ATAATAACAA AGGTATCACG ATTAACTTA ACGCTATAGT CGCAAAATAT GAAGATAAAA	1860
	AAGAAAAAAT TGTAGATGAA ATTGTTTATT ACGTTGATG	1899

55

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3774 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 660:

10	GTATAATATC CTTTCAATCT GTTTTCATAT TTTATATATT TTTTAAATAT AAGTGCTAAA	60
	TGTTTTAACT AAAGCATAGA TTGACAAGAT GTTATACAGA ATTTCAAATT CTATCCAATA	120
15	TTGTTCGAAG TGTAGTATCA CTGGATTGGT ATTAAACAAT GTAAAGGAGA GATTGCAAAT	180
	GCCGTATAAT TACAAGAAAC AAAATGGAGA GTTAATGTCT GTAATGAGCC AAGGTGAAAA	240
	GTTTATTCAT CAATCACCCG TTAATGATGA ACTTAGTGCA TTGATTAGT TATTAATTTT	300
20	TAAAATTAAC GGTGTGTCATT ATTGTGTTGA TATCCATAAA AAAGAATTAA AGGAATTGGG	360
	TGTAAcACAA ATGAAAATTG ATGAAGTCTT GAGTTTTAGA CATTTAGATT TATTTACTGA	420
	TCAAGAAAAA GTGACGCTTG AATTGCAGA AATGTTAAAT TCAATCAAAG ACTTTAAGAA	480
25	GTTTGAAATT ATTGACCGGC TAAAATCATT TTATGATGAA GAACAAATTA TTGATCTTGT	540
	CTTTGTGTGA AACCAAATTA ACGGTGGA CAGATTAAAT ATTATTAGTG ATAGACTATA	600
	ATTGTTTCATA TAAATGCAGA GTTTCATCTC GAACGCTATA TCATAACAAA TCATGCCACT	660
30	ATACAGGTCA AATCTTGTAT AGTGGCATT TAAATTTATCC CTTTGAATAC TGTTATTTAA	720
	CGAATATCGG TCCACCTGGT CCAACAGGGA TACCTAATAG GAACCAAATG ATGACAAATA	780
35	CTGTCCATAC AATACTTAGC GCGATTGAAT ACGGCATTAA ACTAGAAAGT AAGGCTCCGA	840
	GTTTCATGCG TTTATCGTAT TTTTGTGCAT AAGTTAATAA TAAAGGTAAG TACGGCATCA	900
	TCGGTGTAAT TGGATTGGTA ATTGAATCGC CTACACGGTA AATGACTTGT GTGAATGCGG	960
40	GATGAAAGCC GATAAGGATT AACATTGGTA CGAATATCGG TCCTAAAATA CCCCATTAG	1020
	CCGATGCGCT TCCGATTAAAC ATGTTGACCA TTGCACTCAG TACAATAATA CCTAGTATCA	1080
	ATACAATACC GTTTTGATGT TCTAATAATT TGGCACCTTT AACAGCAGCG ATAATTCCTA	1140
45	AATTACTCCA CTTTAAATAC GCAAGTAGCT GTGCTGCAAA AAACACAATA ACGATAAATG	1200
	TTCCCATGTA TCCTACAGCA TCGCCGAACA TTTTACCTAA GTCTTTTGTA TTTTAAATTT	1260
	CTTTGCTTAA AATCCCATAA ACTAATCCAG GACTAAAAA TACGACAAGA ATAATTAATC	1320
50	CGACACCGTT AATTAATGGC GCATCGTCTA GTAAGCTGCC TGTTTTAGCA TTTCTTAAAA	1380
	AGCTATGTTC AGGAATGGCT GTAATAATTA ATAAAATAAT TGTGACTATG AAAGTATAT	1440

55

## EP 0 786 519 A2

	CATCATGCAT TAAACTGTCA TCATATTTTC CTAATCTAGG AATAATGAGC TTAGTTGTAA	1560
	CTAGCAATAT CGTAGGAAGT AATACAACGA CACTCGCTGC GATAAAGTAC CAGTTCATAG	1620
5	CAACGTTTGT TTTAATAGAA TCTGAAACGA TACGTGTTGC CGGTTCTGTA AATGAATAGA	1680
	CCAAAGCATC TTGCATACCA ACAACTATAT TTGCTGCAAA TCCTCCAACA GCGGAAGCAT	1740
	ATGCCATCGT TAGTCCAGCG ATAGGGTGAT AGCCAATTTT AATAAAAAGC ATTGCTGCAA	1800
10	GCGGCGGCAA GATAATTGTC GCAGCATCGC CGGCTGTACT ACCTAAAATA CCAATTAATA	1860
	TAATAGTCGG TAAAATTAAG AAACGTGGTG CGCGATTAC AACAGAAATC ATTAACCTAT	1920
15	CGAAGTATCC TGTTTTCTCT GCAACACCAA TACCAATCAT CACTGCTAGT ACTAAGCCTA	1980
	ATGCTGGGAA CTCTGAGAAA TTTTAAATCG TATCATTAT TATCATCGTA AATCCATCAT	2040
	GGCTAATTAT ATTTTAAATA TAAATGGTTT GATGCGTACC TGGATGCTTA ACAGATACAT	2100
20	TAAATAATGA GATAACCCAT GTCATAATGG CTAAGCCCAC ACACATTAAA AAGAATAAGA	2160
	CGCTAGGATC TGGCAATTTA TTTCCGATTT TTTCAACACT ATTCAAGAAA CGATTGACGA	2220
	TAGACCCCTT TTGTTGATGT TTTGATGTCA TCAATTATTC CCCCCTTTGT TAAATATTTA	2280
25	AAGTGTAACA AAAAATACTC TCAAAAGTAA CAATTTTCAG GAAATAAAAA AACTAATATT	2340
	GTTAAATATT TTGAGTTATT CAATAGAAAG TGTATAGCAG AGTAGTTAAG ACTGCCTGAA	2400
	GACTTATCTA TTAGGTTTAT GAAGCATCGA ACAGTGGAAA ATAAGGACTG TAAGTTTAAG	2460
30	ATATGTTGTA TAGGAGTGAC TGAATGAAAC GTTTGGAAAA TAAAGTAGCA GTCGTAACAG	2520
	GAGCAAGTAC AGGTATCGGT CAAGCTTCTG CAATCGCTTT AGCTCAATAA GGTGCGTATG	2580
35	TATTGGCGGT AGACATAGCT GAAGCGGTAT CGGAGACTGT CGATAAAATT AAAAGTAATG	2640
	GTGACAATGC GAAGGCGTAT AATGTGGATA ATGCAAGCGA ACAACAAGTG GTAGACTTTG	2700
	TGTCTGACAT AAAGGAACAG TTTGGAAGAA TCGATGTGTT GTTTAATAAT GCCGGTGTGG	2760
40	ATAATGCGGC TGATAGAATT CATGAGTATC CAATAGATGT GTATGACAAG ATTATGAATG	2820
	TAGATATGCG TGGGACATTT TTAATGACGA AAATGATGTT ACCTTTAATG ATGAATCAAG	2880
	GTGGCTCTAT TGTTAATACG TCTTCATTTT CCGGACAAGC AGCAGACTTG TATCGCTCTG	2940
45	GATATAATGC TGCGAAAGGT GCAGTGATTA ATTTTACAAA ATCAATCGCA ATTGAGTATG	3000
	GCCGTGATAG CATTCGAGCC AATGCGATTG CACCAGGTAC AATTGAAACC CCGTTAGTAG	3060
	ATAAACTGAC AGGTACGAGT GAGGATGATG CAGGTAAAAC ATTTAGAGAA AATCAAAAAT	3120
50	GGATGACTCC GCTGGGACGT TTAGGTAAAC CAGAAGAAGT TGCTAAATTA GTAGTCTTCT	3180
	TAGCATCTGA CGACAGTTCA TTCATAACTG GAGAGACGAT TCGAATTGAT GGTGGTGTGA	3240
55		

AATTATTTAA AATCGATATC CAGTGGAAAA GAATTTGGCA TGAAGTAGGC AATGTATGCA 3360  
 TATTGTCGCA ATGAAAGACA TTAAAGCGG AGATTAAATTA ATCTTTGAAA ATCACATATT 3420  
 5 GTTCATTGA AGTGATTGC TTAAAGCTTG TTTAACGTA TTGTAGGTCG TGCATCTAAT 3480  
 TTATCAGAAT CTCGAGTTCA AACTTTTGGG GTATCTTTAA AATAAGCTGT ATTTTGTCTA 3540  
 10 TTCTAATAAA TTAAGGAGAA TTTTATGTTA AAAGAAAAAG AAAGTTTGTAG ATTGCTATAT 3600  
 CAAGCTATAA GAGAGATTGC AGATAAAATT GGAGATAATC AGTTAGAAAC TAATCCGTT 3660  
 AGTTTATtAT tATTGGACTT TGATTTTGAA CATGAAGTAT TTGATGAATT GTATCTGTG 3720  
 15 ATTTtAAAAT ATTAAATAC AGTAAGTATA GAGAACATAA GTCATAGTGA GCTT 3774

## (2) INFORMATION FOR SEQ ID NO: 661:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1078 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 661:

TACTGGTTTT GGATTTTGGG GGTAATACAG TACCTAATAG TAATtATGGT GTCGTCaAA 60  
 30 TTATATCTTT CCGTGATGTT CTTACTCATT AGAACATCnG CCTTCAGAGG AATCATGATA 120  
 CGAGGAATAA GAAATTTAAA TGTGAGCGAA GTCAATATAG TATTTGCGAT TATTTTtATT 180  
 AACTCCATTA TTATGTtTAG TTTGATTTTT CGAGGATAAC TTCAATTTTT GCATTTTGAG 240  
 35 GTTTTTTAAAC ATATCTATTT GCATCAGTTG ATGGCAACCT TTTACTTAAA TCTATTGTGT 300  
 AGTTATtGTC TGCACCTGTT ATTTTAATTT GTCCTTTTATT ATAAGAATTA TTATATAATT 360  
 TTTTACTTTTT AATTAAATGTT TGACGAATAC GAAAATCTAA TTCTTTTAAA GTTAAAACAG 420  
 40 GCTTATTGCC TTCATAAACT GGAAATCCGC CAGTAAACGT TTCTGCTTTA TCTTTATATG 480  
 TTACATTtCAG TTTATAGTGT TTATCGTTAG ATGTTGCTGC AGGAGTAACA CCACCAGTAA 540  
 ACGTTTCTTG AGATAATGCA AAAGaATCAA TGGTTTCTTG GTCTTTTATG CyAAAAATAT 600  
 45 CAACGCTTTT ATTTCTTAAT TGGTTGATAT TGCCCCAACT TTCAGGTCCA TAAACTTGAA 660  
 TATGACTATA CCAAGaAAAC TGTAACAACG TTGCATGAAT CgTACCGTTA TCTTTTGCC 720  
 ATAACGTACT GTTAGAGAAG GTTAAATATT TTTGCGAGTA ATATTTAGTT AACTCATTAA 780  
 50 CGTTAGTTTC GTTTTGATTT ATATAAtAAG CTTTCGCTTC AGATGAAGAA TTGATAkGTG 840  
 TATTAGGAAA TTGTGTAGAT GCTGTACCTA ATAGTAACAA TGTTGTtGAT AAAATAATTT 900

55



ATAAAAAGGG GTTAATTAGA TAATTGAAAT TATCCGCATT TACAAAAGGT AATAGGTTAG 1020  
 TTAGATTTTT CGAGTATGAC TcAATTTCTG CATTACGAGG ATTTTAAACA TAACGGTT 1078

5 (2) INFORMATION FOR SEQ ID NO: 662:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1398 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 662:

AACCTGCTTT TTAGTATCTA CGTTAATATT AATAGCTTTA ATTTCACTTG TATTAATTAA 60  
 ATCAGCTGTG TAAATACCTG CTTTCAAATC GATAACTTTC TTGTTCCAT TTTTAAAGTA 120  
 20 AACAGTATAT TTCGCTTGCT TCGATAGTCT TAAATCTATA TCACTAATAC CTCTGTCTGA 180  
 TTTTAAACT GATTAACTC TATCCTCTAA ATCTTTATAA CTAATATTTT GATTCTTATT 240  
 AAATGTTAAG CTTGATAAAA TATTTTGGCT TGTACCGTTC ACAGTGaTTG CATATGGAAC 300  
 25 ATGGACTTTA GAATATCCAT GGTGTAACGA ACTTGATGAT TTATCTAATG GCTTAGCTGC 360  
 GGCAGACGCT TCATTATTAT TAAAGTTTGC ACCTGTTGAT GCTAAACAC CTAATGCTAA 420  
 AGTTGTTGTA ATCAATGACT TAAATTTTCAAT AAATTATCTC TCCTTTTTTG TGTAATTCGT 480  
 30 ATTTGCAACT TAATTATAGC CAGACTTTCT CTATTTTTTG AATTAAGTGA ATATTAATAA 540  
 TAAATTATCT TTAACAATAA TTTTAAACA CTGTTAAAG TTCTTTTAAT TTTGATTAAAC 600  
 TAATTAATTT ACAATACCTA AAATGTTGTT TGGTTTTGTT TATACCAAGC TTCAAACCTA 660  
 35 AATGTCATAA CAACATTCAT TTCTTAATTC CTATTAGATT TGTCGATTAT ATTTACAGCA 720  
 TCTTTATACT CAAAAACAT TTAATTAAAA ATATAAATTC GATTTAATAA TTAATTTAAA 780  
 40 TTTAGTTAAT CAATTTTGCA TCTATTTTGT TGTAAGCTAT ATAAAAGGAG TGATAATGAT 840  
 GGTGAAAAAA ACAAATCCA ATTCACTAAA AAAAGTTGCA ACACTTGCAT TAGCAAATTT 900  
 ATTATTAGTT GGTGCACTTA CTGaCAATAG TGCCAAAGCC GAATCTAAGA AAGATGATAC 960  
 45 TGATTTGAAG TTAGTTAGTC ATAACGTTTA TATGTTATCG ACCGTTTTGT ATCCAAACTG 1020  
 GGGGCAATAT AAACGCGCTG ATTTAATCGG ACAATCTTCT TATATTAAAA ATAATGATGT 1080  
 CGTAATATTC AATGAAGCAT TTGATAATGG TGCATCAGAC AAATTATTAA GTAATGTGAA 1140  
 50 AAAAGAATAT CCTTATCAAA CACCTGTACT CGGCCGTTCT CAATCAGGGT GGGACAAAAC 1200  
 TGAAGGTAGC TACTCATCAA CTGTTGCAGA AGATGGTGGC GTAGCGATTG TAAGTAAATA 1260

55

CAACAAAGGC TTGTTTATA CAAAAATAGA GAAAAATGGT AAGAACGTT CACGTTATCGG 1380  
TACACATACA CAATCTGA 1398

5

(2) INFORMATION FOR SEQ ID NO: 663:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1200 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 663:

15

TTGTAATTGG AACGTACGAA CTTTTCTAGG TAAGAAACGT CGAATCTCGT CCTCATTATA 60  
ACCAACTTGT AGTCGTTTAT TATCTAAAAT AATTGGACGA CGTAATAAGC CAGGATTATC 120  
TTGAATGATT GAATATAAGT CTGTGAATGG TAGTGAATCA ATATCAACAT TTAATTTTTG 180  
GTATGTTTTA GAACGTGTAG AAATGATTTT ATCAGTACCG TCTTCAGTCA TTTTAAATAT 240  
TTGCTTAATT TCATCAATTG TTAAATGTTT AGAAAAATA TTACGCTCCG TATACGGAAT 300  
GTCCATGTTT tTGkTAACCA TGCTTTTCGCT TTTACGGcAA GATGTGcAaC TTGGtGaAGT 360  
aAATAATGtT ACCATACATC TCACTCTCCT ATTTGAATGA ATAAAATTCA TTGCTTAAAA 420  
TTTAGTTATA GATCAAGAAA AAACATTTTT TTCTAAAATT CTTAATCGTT ACTATTTATT 480  
ATAACTATCT AACATTAAAA TTAAATGAGA AAAACCTAAT TTTTCAGATA AGTTTCTACA 540  
CTTATAAAAA AGATTATTA TCCCTTTGTT AGTAGTAAGT TATACGTATA TTCTAACACA 600  
TCTTACATTT TTAAGAAATA CTGTTATAAT GATAATTATT AAAATATTAC TAAGAAAGTA 660  
GGCATTTAAA TGGAGACATT ATTTTCAGGC ATCCAACCTA GTGGAATTCC TACTATTGGA 720  
AATTATATTG GCGCACTAAA ACAATTTGTT GATGTGCAAA ATGACTATGA TTGTTATTTT 780  
TGTATCGTAG ATCAACATGC AATTACAATG CCACAAGATC GTTTAAAATT ACGTAAACAG 840  
ACCAGACAAT TAGCAGCGAT TTATTTAGCT TCTGGTATAG ATCCAGACAA AGCAACATTG 900  
TTCATACAAT CTGAAGTCCc TGCACACGTA CAAGCAGGAT GGATGTTAAC TACGATTGCT 960  
TCTGTTGGAG AATTAGAGCG TATGACGCAA TACAAAGATA AAGCTCAGAA AGCAGTTGAA 1020  
GGTATACCTG CTGGTCTATT AACATATCCA CCTTTAATGG CAGCTGATAT TGTTCTTTAC 1080  
AATACTAATA TCGTTCCAGT TGGAGATGAC CAAAAGCAGC ATATCGAATT GACTCGTAAC 1140  
CTTGTAGATA GATTTAATAG TCGCTATAAT GATGTGCTTG TGAACCTGAA ATTCGTATGC 1200

50

(2) INFORMATION FOR SEQ ID NO: 664:

55

(A) LENGTH: 787 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 664:

10	TGTACCTTTT GCTTTAATAA ATACTGTTTC TTTATCATAA TTAGGTTTCAG TTAAGAAATT	60
	AAATTGTAAG CTTTGAGTAA TATTTTTTTG ACTATCACTT GTTGTAGCTG TACGTGTATA	120
	CATTTTAGTA TCACCATCAA GATTCTTCTC AGAAACTTGT TTGATTTCAG AATTAATCTT	180
15	TGCAAATGAA GTAGCTGGTA ATACAGTGAA AGTAGTCGAT AGTGCTAAsG wACAAATTGT	240
	GATATTTTTA CATAGTTGTT TAATCATTAG TAATCCGCCC TTTCAATATT ATCCTTCTTT	300
	ATAAGGTTTA TTGTCATCAG AATATTTATC AACGACTTTA ACTGTTTTAT TTTCCAATC	360
20	AACTTCATAA GTGACAATTA ATCTTTGACC ATCTTTATTT TtctCTAAAA TTGGAGGTGC	420
	ATAATGTATy CCAGGTCTGT TTTTCAAAT ATCTTGATTT CGTGTGTATG TTACTTCAAA	480
25	TTGCGTTTTC TCATTTGACT TTTCATTAGA TAAATAAGTT AAAAATTCTG GATTAAAGCC	540
	ACTTCTTACT AATGCTGGGT ATCTATATTT TGAAGCAAAG CTTAGTTCaG GGTTTTCTAC	600
	AGTAGCAATT CTCGTATTTT TATAGAATAA TAATTCATCA TTTCTATTTT TCACTTCTCC	660
30	ACCATACTTC AAGTCATTCG CAATAACTGA CCAGTGACa TGgCCAGTTn ATTTATTTTT	720
	ACCGCTGGCA ATTGTGTCAT AATTGCTGC ATCATCAACT AATCGTTTTG GAAGTAGCTA	780
	TTTGCTG	787

35

(2) INFORMATION FOR SEQ ID NO: 665:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 533 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 665:

45

	TTATCTGTCA TTAACCACTG GTACATGTGA ACCCGGChTn AATCAATATA TATATTTAAA	60
	AGCAAAGGCG CGCCATGTGC CTTTtTTnTA tTTTGATATAT CTGTATCAAA ATCGATTTGA	120
50	TTAAAATCCG CTTTATTTTA TCATCTATTC AAATGATTTT AGTGCGCTTA TTTTACTATG	180
	GCATTATTGC CTAACCTGTT TGAGATATAT TAATATTTGA TGATTCGTAT TCAAAGTTTT	240
55	CATTTAACAT GTATTTAGTA TCATGATAGC TGCTTCATTG ATGATATACT ACCTACTTTT	300

CAGAGGCAAT ATTGAACAAT TCTTCATCAT GACTATTTTC CATCACATAA CTATGCTTAG 420  
 CGAACGCTAA CATATCTTTA TCATTATTCG CATCTCCGAA GGCCATGAGC TCTGAAGGAG 480  
 5 ACATTTCCCA TTTATCTAAC AATCGTTTTA ATGCCTGnCC TTTAGTCATG TTT 533

## (2) INFORMATION FOR SEQ ID NO: 666:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 666:

GACTTCTCTT GCACTAAAAA TTGCTCAAAT TGCTCGCTGT CATAACCAGG CGGTGTTTCC 60  
 20 AACCATACAT ATATGCCACC TTTAGCATGA ACAAATGGCA AATCAGCTTT TGCAAGCATG 120  
 GCTTCGAATC GGTCACGACG TGTTTTAAAT ACATTGCTTT GTTCTTCTAA AAAATCATCA 180  
 25 TAATGGATTG AAAGCATATA TTGCGGTCAT CTTGTAAATG CACCAAACAT CCCAGCATTT 240  
 GTGTGCGTTT GGTACTTTTT CAAAGCTTGA AATCATATCT TTATTACCAA CTGCCAAAAC 300  
 CGACTCTGAA AACCTGGACA TGTTGATGAA CCTTTnAGAC AAGGGAAGAA ATTTGCAATC 360  
 30 GCAACATCTT TGCCCATTTT nCCGAAGCAA GTGnGACTAG 400

## (2) INFORMATION FOR SEQ ID NO: 667:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 412 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 667:

TCTCCAATGC TACTCAAATT AAAAAGGTTT TAAATATTGT TAATTCTGAT CCAGAGCGAA 60  
 45 AGATTGTTAT CGTTTCTGCT CCAGGTAAAA GACATGATAA TGATATTAAA ACAACTGATT 120  
 TGTTAATCAG ATTATATGAA AAGGTCATTA ATCATCTTGA TTATCATGAT AAAAAAAGAG 180  
 AAATTATTCA GCGTTATGAT GATATTGTAA AAGAATTGCA AATGGATGAA AGTATTTTAC 240  
 50 GGACGATAGA TGTGACTTTG GAACATTATA TAAATCAATT AAAAAATGAA CCAAAGAGAC 300  
 TATTAGATGC ATTACTTTCT TGTGGTGAAG ATTTTAATGC GCaA<sub>y</sub>TGatA GCyTTATATA 360  
 ATAatAGTCa gGTtACCAAC AAmATwTATA TCCcCGAAGG AAGCGGGTAT TT 412  
 55

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 668:

TTACTCGTTA AAGATATTGT AAAAGATGAA GTGACAGAAT ATGACATTCA TCAAATGTTA	60
CCGCATCCGA TTAATATGGT AAGGGTTAGA CTTTTTGGTG TGAAATTAAA AGAGATTATA	120
GCTAAAAGTA ATAAACAAGA nTATATGTAT GAACATGCAC AAGGTTTGGG TTTCAGAGGG	180
AATATATTTG GAGGATATAT TCTTTATAAT TTAGGGTACA TTCATTCTAC AGGGCGTTAC	240
TATCTGAATG GAGAAGAAAT CGAGGACGAC AAGGAATATG TACTAGGTAC GATAGATATG	300
TATACGTTTCG GTCnTATTnC CCAACATTGA AGGATTACCA AAGAGTATTT AATGCCAGAG	360
TTTTTCAGAG GTATATTTAA nGAAAATTAT TGGCTATTTA	400

(2) INFORMATION FOR SEQ ID NO: 669:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1236 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 669:

TGTTTCGATAT TTTTAAATTT ATCTTTTAAA TACAACAAC TTTCCGTAA TGATTTAACT	60
GTTTTATGAC TAATGCCATT GAATATTTCT AGCGTTTTAT TTAAC TTATC GATAATCGCA	120
TGTAAATCCT TCAAAATGTC TTTTGTTC AAGTAAATA CATTATGGAA GCGATGAATA	180
TCATCATCAT AAACATCAGA ATCATTGATA ATCGTAAATA TCGTTGAGAA CAATTGCTCA	240
TTTAACTCAT GAATCTCATT CATACTAGCC TTCAAGCCAA AAATATCAAT TGGTGCAATA	300
TCTAATTTTT CCAAAATTCG CTGCTTTTCC AGTTGATCAA TTGCCTTTAA CAATTTTTCA	360
TTTTCGTTTT TACCAATCAA ACCAAGCTGA TATTTAATAT CAGCATAACT CAACTCATTT	420
GTCAC TTGAT TTAAGGCATA GTCTGGTAAG CGATGTGCTT CATCCACTAT ACAATCATCA	480
AACAATTGAT ATATTGAATT TTCAACATCA GAATGaaTTA AATGTGCATG ATTTGTAATA	540
CCAATTTGaa TGTTCTGTGC ATTTTCGCTTA ATAAAATTAT aATAATGAAC ATCGTGACGT	600
CCCGGTACAT ATGTTTCAAT TTTCTGGTca AAATACATCT TTTGACCACC TTTTAAATTT	660

ATATTCACtT CGTAATTACT TGTGTCATCT TTTAAAATTT GACTAATAAG CCCCAATGAA 780  
 ATGTAATCaC TTtTACTTTT AATCAATAGT GCATTAATTT TAAAATTCAA CGCTTCATTG 840  
 5 ATTGCTGGAA TATCTTTTTC TAACAATTGA CTTTGCAGTA ATTTAGTATT GGTAGAAATC 900  
 ATGACATGCT TCCCAGTTTC AATATTATAC ATCAAGGCCG CAAGTAAATA TGCTAATGAT 960  
 10 TTACCACTGC CTAGTGATGC TTCAATCATT GCTTTTTTCAC TATGCATGAG CTGATCTAAT 1020  
 ATAGTTTCCG CTAAATATAA TTGTTGCGGT CGATATGTTA AGCCAAGTTG ATCTACAGCT 1080  
 TTGCTATATA AAGACTTCAA GCTGCCATTA TAATTTGTTG TCGGCTTTTT AAAATCAACT 1140  
 15 TGCTTACGAT AGATAATCTG TTCGAACTTT TCGTACGATT TATCCAATGG CTTTGCATCA 1200  
 TATTGCCTAA CCATCTCAA GAAAATATCA TACAAA 1236

## (2) INFORMATION FOR SEQ ID NO: 670:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1819 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 670:

30 ACaSaATTAT TGACCAATAT GACTCGTGGA CTGATATGTT TAAAGCACTA CTGCATGAAA 60  
 CATTTAAGC ATATGGCGTT CTATTTATAG ATGCGCAGTT TGAGCCGTTA AGAAAAATGG 120  
 AAGCGCCTAT GTTTAAAAAG ATTTTGAAAA AACATCAGTT GCTTGATGAT GCTTTTAGAG 180  
 35 CAACACAACA ACGTACTCAA AATCAAGGCT TGAATGCGAT GATACAAACA GATACAAATG 240  
 TTCATTTATT CTTACATGAT GAAAATATGC GTCAATTAGT tTCGTATGAT GGTAAGCAyT 300  
 TTAmATTAAA TAAAACAGAT AAGACATATA TAAAGGAAGA AATTATAAAT ATTGCGGAAA 360  
 40 ATCAACCKGA ATTATTTTCT aATAATGTAG TGACAAGACC ATTAATGGnA GAATGGTTAT 420  
 TTGAACACGG TGGCATTGTG TGGAGGACCG AGTGAAATTA AGTaCTGGGC TGAACATAAA 480  
 45 GATGTATTTG AACTATTTGA TGTTGAAATG CcAtATCGTA TGCCAAGGCT TAGAATTACT 540  
 TATTTAAATG ACCGTATAGA AAAATTACTT TCGAAATACA ATATTCCATT AGAAAAAGTG 600  
 TTAGTCGATG GTGTTGAAGG AGAAAGAAGT AAGTTTATTA GAGAACAAGC ATCACATCAA 660  
 50 TTTATTGAAA AGGTAGAAGG TATGATTGAA CAACAGCGTC GTCTAAACAA AGACTTATTA 720  
 GATGAAGTGG CGGGGAATCA AAATAATATT AACCTTGTGA ATAAAAATAA TGAAATTCAT 780  
 ATACAACAGT ATGATTATTT GTTAAAACGT TATCTTTTAA ACATTGAAAG AGrAAACGAC 840

55

5 GAAAGAATAT GGAATCCACT TCAAATTTTG AATGATTTTG GGACAGATGT GTTCAAGCCC 960  
 TCCACCTATC CACCACTTTC TTACACTTTT GATCATATTA TTATAAAACC TTAATATACC 1020  
 AAGGGTTTAG CCCGATTTAT CTTAATGATA AATCGGGCAT TTTTGTGTTT TTTAAAATAA 1080  
 ATTTACAAA TTTTGTATAA ATAGTGGTGG ATAGTGGGGA GATGTGGTAA ATTATATATA 1140  
 10 AGGTGAGGTG ATAAAAATG TTCATGGGAG aATACGATCA TCAATTAGAT AAAAAAGGAC 1200  
 GTATGATTAT ACCGTCCAAG TTTGTTATG ACTTAAATGA GCGTTTTATT ATCACAAGAG 1260  
 GCCTTGATAA ATGTTTATTC GGTTACACTC TAGACGAATG GCAACAGATT GAAGAGAAAA 1320  
 15 TGAAAACCTT ACCTATGACA AAAAAAGACG CACGTAAGTT TATGCGTATG TTCTTCTCTG 1380  
 GTGCTGTTGA AGTAGAACTT GATAAGCAAG GCGTATTAA CATCCCTCAA AACTTGAGGA 1440  
 AATACGCTAA TTAACTAAA GAATGTACAG TAATCGGTGT TTCAAATCGT ATTGAGATTT 1500  
 20 GGGATAGAGA AACTTGGAAT GATTTCATG AAGAATCTGA AGAAAGTTTC GAAGATATTG 1560  
 CTGAAGATTT AATAGATTTT GATTTTAAA ATGGAGGAAT TGAAGtGTTT CATCATATCA 1620  
 GCGTTATGTT AAACGAAACC ATTGATTATT TAAATGTAAA AGAAATGGT GTGTACATTG 1680  
 25 ACTGTACGCT AGGTGGAGCG GGACAnGCCC TTTATTTACT AAATCAATTA AATGACGACG 1740  
 GAAGATTAAT AGCAATCGAT CAAGACCAA CTGCAATTGA TAATGCTAAA nGGGTATTAA 1800  
 30 AGGATCATTT GCATAAAng 1819

## (2) INFORMATION FOR SEQ ID NO: 671:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 609 base pairs  
 35 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 671:

ACCAATTATT GGACAGGAAT TAGCAAAAAA TGCAATGCTT GCATTAATCT ATGCATCGAT 60  
 45 AGGTATCATC ATCTATGTAT CATTACGATT TGAATGGCGC ATGGGTCTTT CATCTGTATT 120  
 GGCATTATTA CATGATGTAT TCATTATAGT AGCGATTTTC AGTTTATTTA GAATTGAAGT 180  
 AGATTTAACA TTTATCGCCG CTGTATTAAAC AATTGTCCGT TATTCAATTA ATGATACAAT 240  
 50 CGTAACGTTT GACCGTGTAC GTGAAAACCT ACAAAGGTT AAAGTGATTA CGACAACAGA 300  
 ACAAATTGAT GATATCGTTA ATAGATCAAT tAGACAGACA ATGACACGTT CAATTAATAC 360  
 55 AGTATTAACA GTTATTGTAG TAGTAGTTGC TATACTATTC TTCGGTGCTC CTACGATATT 420

TGCCGTTCCG CTATGGGGAA TAATGAAAAA ACGTCAGTTG AAAAAATCGC CGAAACACAA 540  
 ATTAGTTGTA TATAAAgAAA AGAAATCGAA CGATGAAAAG ATTTTAGTTT AAAAtGaATT 600  
 5 AAGCGGTAT 609

(2) INFORMATION FOR SEQ ID NO: 672:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2222 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 672:

CTTAAAACAG CAATTTGAAC GTTTACAAAA TGAACAAATC TTTGTTTATG TTTGTCATGG 60  
 20 TAATCACGAT CCTTTATCAT CAAAGATTTC ATCAAACCTGG CCAGATAATG TtnTTGTATT 120  
 TTCAAATAAA GTTGAGACGT ATGAAGCAAT TACTAAATCT GGTGAAACAA TTTATATTCA 180  
 CGGATTTAGT TATGAAAATA GAGCAAGTTA TGAGAACAAG ATTGATGAAT ATCCATCAAG 240  
 25 TCAAGGCCAA AAAGGCATAC ATATTGGTGT CTTGCATGGT ACGTATAGTA AATCTTCAGT 300  
 TAACGAAAGA TATACCGAGT TCATTTTAGA AGATTTAAAC AGTAAATTGT ATCATTATTG 360  
 GGCTTTAGGT CATATACATG AACGTCAACA ATTAAGTGAT ATGCCTGTAA TTAACCTATTC 420  
 30 AGGTAATATT CAAGGTAGAC ATTTTAATGA GCAaGGTGAA aAAGGTTGCT TATTAATCGm 480  
 GGGTGACCAC TTAATAATTAA AGACTAAATT TTATCCTACA CAGTATATTA GATTGAAGA 540  
 35 AGCAACTATT GAAACGGATA AGACATCTAA GCAAGGTTTA TACGAGGTCA TTCAAACTT 600  
 TAAAGAACAA GTGAGAGAAG AAGGAAAAGC CTTTTATCGT TTAACGCTTG TTATTAATAG 660  
 TGAGACATTA ATTTACCTC AAGATTTATT ACAAGTTGAA GAAATGATTA CAGATTATGA 720  
 40 AGAAAACGAA AATCAATTTG TATATATTGA TGAGTTAAAA ATACAATATG CACAAAATGA 780  
 TGAGTCACCT TTAGTTAATG AATTTTCAGC GGAATTATTA GTCGATCAAA CTGTTTTTGA 840  
 TAAAGCGATG TCAGATTTAT ATTTAAATCC AAGGGCATCT AAGTTCCTAG ACGATTATGG 900  
 45 AACATTCGAC CATAACGAT TAGTTAATCG TGCTGAAGAA ATATTAAAAG CTGAAATGAG 960  
 AAGTGAACAA AATGATAATT AAATCACTTG AAATTTATGG TTACGGTCAA TTTGTTCAaC 1020  
 50 GTAAAATTGa ATTTAATAAA AAcyTCaCTG AAA1TTTTGG TGAAAATGAA GCGGGTAAAT 1080  
 CGACGATTCA AGCATtCATC CATTGATAT TATTTGGATT TCCAACATAA AAGTCTAAAG 1140  
 AGCCAAGACT AGAACCACGT CTAGGTAACC AATACGGTGG TAAATTAGTA CTTATTCTTG 1200



TATATTTACC TAATGGTGCT GTGCGTGATG ATGCTTGTTT AAAAAAGAAA CTTAATTATA 1320  
 TTTCTAAAAA GACATATCAA GGTATCTTTT CATTTGATGT ACTAGGGCTT CAAGACATTC 1380  
 5 ATAGAAATCT AAATGAAAAA CAATTGCAAG ATTATTTATT ACAAGCmGGG GCTTTAGGAT 1440  
 CAACTGAaTT CACGTCAATG CGCGAAGTGA TTAATCGTAA AAaAGATGAA TTATATAAAA 1500  
 aATCAGGTAA AAATCCGATC ATTAATCAAC AAATTGAGCA ATTAAAACAA CTAGAAAGTC 1560  
 10 AAATTCGTGA AGAAGAAGCA AAGCTAGAAA CATATCATCG CTTAGTAGAT GATCGAGATA 1620  
 AATCATCAGC TCGATTAGAG AATTFAAAGC ATAATTTAAA TCAATTATCA AAAATGCATG 1680  
 15 AAGAAAAACA AAAAGAGGTT GCTTTACATG ATCATTCAACA AGAATGGAAG TCTCTAGAAC 1740  
 AACAGTTAAA TATTGAGCCA ATCACATTCC CAGAAAAAGG TGTGGATCGT TACGAAAAAG 1800  
 CACGAGCGCA TAAGCAATCG TTAGwAAGAG ATATTGGTTT AAGAAATGAG CGTTTAGCTC 1860  
 20 AACTTAAAGA AGAAGCGACT CAATTAGAGC CAGTTAAACA ATCTGATATT GACGCCTTCA 1920  
 TTAGTTTGAA TCAACAAGAA AATGAAATTA AAAATAAAGA ATTTGAACTT ACTGCAATCG 1980  
 AAAAGGATAT TGCGAATAAA CAACGTGATA AAGATGAATT GCAATCAAAT ATTGGTTGGT 2040  
 25 CTGAAACGCA TCATGACGTA GATAGTTCAG AGGCAATGAA AAGTTATGTC AGTGAGCAAA 2100  
 TCAAGAATAA ACAAGAACAA GCTGCATACA TTAAACAATT AGAACGTAGT TTAGAAGAAA 2160  
 ATAAAAATCGA AGATAATGCG GTTCATAGCG AACTAGATTC TGTGAAGAA AAATAGTTCC 2220  
 30 TG 2222

## (2) INFORMATION FOR SEQ ID NO: 673:

- 35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 406 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 673:

45 TTGCAGGTAT CATTTTTTTA ATGCCATATG GATTATGTTT TCTACCGTTT TATAAGCAAA 60  
 AAAAGAAAAA ACAGACATTT AAAAAATACA TGGTTTACAC TACGATTGGT TTGTCAATTT 120  
 GTCTAGGCTT ATCTCTAGTT TTGGTTCACA CTACGAAAAT TTATATGGAC GAAGGTGGCG 180  
 50 TAAGATACTA TTACGGTAGT TTTGTAATGA AACAAAGCGG CGTTATGCT TATTTAGCTT 240  
 TAGCGGTACT TTCAACGTTG TTAATTGTTG CGAAAAAGC TACAAATAAA AATAAAGAAA 300  
 TCGAAACCGT CGACAATACA AATATAACGG AAAGATAATT AAGGGAGTGC TCATTGAGGA 360

55

## (2) INFORMATION FOR SEQ ID NO: 674:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 927 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 674:

```

CCACTTTGTG GATTTTCTTT ATATTCTCCA CGCTTGtACT ACAACTTCTT TCGTTTCTGT      60
TTCATCACCG ACTGCCGCAT TCGTTAGCAC ATGTAATAAC TCTTTTGCGG TTAATACATT      120
CTCATCTATA ATCTTATCTT TTTGTTCTTG TATATATTGC TTGATGTGCG GCTTTTTTCAA      180
TAACCTACAC GCTGTCACAT GTGCGCTATT TCGGTTATAT CCTGyTTTTA TGGCACTTTG      240
TGTTACATTC AGTGTTCCTAA TATACTCATT CACAAAACGT GCTTGCTTTG CAGTTAACTC      300
ACTCATTTTA TCACCCCCAC AATTTTATCT AATATGGTTT CATACCATAA TATTACAGAT      360
TGTTCTGAAC AATCTAAGGC ACTACTAATA TCTTGATAAC TAAGTCCTTG TATAAGGGAG      420
TCAAAAATAT AAAACTCTTT ATCGGTCGCT AATCTGTCAA CAATCATTTT TATGTGATTC      480
TTTATAATAT GATCATTGAC ATTATCGTCT GTCATCAATT CGTCAGAATC TTCATCACCT      540
ATTAAAAAGA AATCATCAGT ATTTATTTCA TCATCGCCCC GTTAACTAGC TTTGAAGTCT      600
TTAGCACACT TGCATATACC GGCTGTCGTG CTGGCKAGAT ACTAGCATTG AAGTGGTCTG      660
ATATTGATTT TGAACAAC ACGATTAGTA TTAATAAAC ATATTACAAT CCAAATAATA      720
ACAAGAAGAA ATATCAGATA CTTCCCCCTA AAACTGAAAG TTCTATCGGT AAGATTTCCG      780
TTGACCCAAA TGTAATAAAG GtGtTGCgtG ATTATAAGAT AAACGTTCAA AATAATTGGA      840
AAAACGAATT ATATAATGAT AACCATTGTA ATGAAAAAAC TATCAATGTG GATACAAACC      900
ATTATGTCAC GTACCAGCAT ACTCAAA      927

```

## (2) INFORMATION FOR SEQ ID NO: 675:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1100 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 675:

```

GATGTTAAAnA TAAGTACATC ATTTTGACTC CCAAATATTG GCTTTAGACC TTGAAATGCT      60

```

CTCATAATGG CATCAGGTAC cAGGGGTTGG GaCCAGGTGT TAATAACAAC GGTGATGAT 180  
 AATACATTTA TATGCCTCCT ATAAATAAGA TTTCACTATT TTAGCAAATT TTCTGAAAAT 240  
 5 TTAAGAGCCT AAAAGtGAT AACGCTATAT TGTCGAACAA ACGTTATCAC TTAACGATTT 300  
 TATTTGGCCT GAATGATGGG ATAGATTTTT ATGTCCACAT TATTTCTTAC AGCATTTGAA 360  
 ATCATGCAAT TATTATCTGC AATTGTTATC AATTTTGTA ATCGCTTTTC TAATTGTGCT 420  
 10 ATTTGATCAC TTGGAATTG AATTGAGGA TGGTGACAA TTTTGGACAT ACTGAATTTT 480  
 CCGTTATTTA AACAAGCTGT TCCAATCGAT TGTTGTkCAA TTGAAATATC TGTGAACTTT 540  
 GCACGTTCAA GAGTAGCTGC TAATGAGATG ATATAACATG ATGAAGCGGC TGATACTAAC 600  
 ATTTATCGG GATTTGTTC TATACCAACA CCACCTAAAG AAGCAGGTAT AGAAATATTC 660  
 TCTGAAAGTA TGTCGCCTTG AACGTTTCCG ACATTGTTAC GACCACCTTG CCAAGAAGTT 720  
 20 TGGACTTTAA AGTCATGTTG ATGCAATTGC TTTAACCTCC AATATAATTG TGATAGTTTA 780  
 ATTTTAGAAT AGTTTATCAA AGTTTAACAG AAAGGTGACT TATCAATGAC TCTGAATAAA 840  
 CTGAAAGATG AATTACAAAT TGTTTCGCAC CGTGGATTGC CGAGTGATT TCTGAAAAT 900  
 25 ACAATGGTCG GTTATCGAGA GGTAATGGGG CTCAATGTTG CTATGTTAGA AATAGATGTT 960  
 CATTGACCA AAGACCAACA TTTGTTGTG ATACATGaTG AAACAATTGa TAGAACATCG 1020  
 GaTGGtArGG GCGTAwTGc TGaTTACACA TTATCGCAAT TAAAATCATT TGATTTTGGT 1080  
 30 AGTTATAAAG ATGTTGCTTT 1100

(2) INFORMATION FOR SEQ ID NO: 676:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 460 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 676:

45 .ATTAATTTCA TATGGAAATA GTTGAACATA CTGCGCTGTA TAAGCTTCAA CAGTTTGATG 60  
 AATTAACGAT TGATCTTCAA TATATCCGTA GAACAAATCT TCAGTACAAA CTACTTTACC 120  
 TTTATCAGGT TTAATTGCAC CTGCCAACAA TTGACCTACC AACGCTTTGG AAGATTCAGG 180  
 50 TTCACCAATT ATACCTAATG CTTCTCCTTG ATAAATATGT AACTAATAT TGTTTAAATC 240  
 GATATCTTCA GCATCATATC CAAAAGGTAA ATACCATTTT TTATTCTGTT TATTCCTATA 300  
 GTAGTGTGTT ACTTTTAGTA ACTTTAAAC AATTGAACTT CCCATCTATT TTCATCCTTC 360

55

CCCCACGCaA AAATACCTTT TAATCTTnCT ACTTTAAAAAT

460

(2) INFORMATION FOR SEQ ID NO: 677:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1451 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 677:

15	GTCAAGTTTT CGGCTAGATT TTAAACGCTC ATCACTATGT AGAATGAAAC TTTCGAAAAA	60
	CTGTAAATCA TCATAACCTT TTACATAAAC ATAACCTTCG CCACCAATTG CTTGAATTAA	120
	ACATTGGGCG GCCATTTGAA TTTCTAAAGA TTGTTTTTCT AGCCTATTAA AGATACCTAT	180
20	TAGTTGTGTG TTTAAGATTT TTGACATCTT TATCCTCCAA TCTACTTATA AAATATTGTA	240
	ATTAATGACT ACATATTATG CAACGGCTTA AATTGTATAA AAATGTATAC GTTTGCATTT	300
	AGTATAACTA TCGCATTTTT CAAAAAATAC ACATTTAATC TGCAGTATTT CAATGCATTG	360
25	ACGCTATTTT TTTGATATAA TTACTTTGAA AAATACGTGC GTAAGCACTC AAGGAGGAAC	420
	TTTCATGCCT TTAGTTTCAA TGAAAGAAAT GTTAATTGAT GCAAAGAAA ATGGTTATGC	480
30	GGTAGGTCAA TACAATATTA ATAACCTAGA ATTCACTCAA GCAATTTTAG AAGCGTCACA	540
	AGAAGAAAAT GCACCTGTAA TTTTAGGTGT TTCTGAAGGT GCTGCTCGTT ACATGAGCGG	600
	TTTCTACACA ATTGTTAAAA TGGTTGAAGG GTTAATGCAT GACTTAAACA TCACTATTCC	660
35	TGTAGCAATC CATTTAGACC ATGGTTCAAG CTTTGAAAAA TGTAAGAAG CTATCGATGC	720
	TGGTTTCACA TCAGTAATGA TCGATGCTTC ACACAGCCCA TTCGAAGAAA ACGTAGCAAC	780
	AACTAAAAAA GTTGTGAAT ACGCTCATGA AAAAGGTGTT TCTGTAGAAG CTGAATTAGG	840
40	TACTGTTGGT GGACaAGAAG ATGATGTTGT AGCAGACGGC ATCATTATG CTGATCCTAA	900
	AGAATGTCAA GAACTAGTTG AAAAACTGG TATTGaTGCA TTAGCGCCAc ATTAGGTTCA	960
	GTTTCATGGTC CATACAAAGG TGAACCAAAA TTAGGATTTA AAGAAATGGA AGAAATCGGT	1020
45	TTATCTACAG GTTTACCATT AGTATTACAC GGTGGTACTG GTATCCCGAC TAAAGATATC	1080
	CAAAAAGCAA TTCCATTTGG TACAGCTAAA ATTAACGTAA AACTGAAAA CCAAATCGCT	1140
50	TCaGCAAAAG CAGTTCGTGA CGTTTTAAAT AACGACAAAG AAGTTTACGA TCCTCGTAAA	1200
	TACTTAGGAC CTGCACGTGa AGCCATCAAA GAAACmGTTA AAGGtAAAAT TAAAGAGTTC	1260
55	GGTACTTCTA ACCGCGCTAA ATAATTAATA TTAGTCTTT AAGTTATTAA TAACGTAGGG	1320

AATAAATAAA ACAGTTTGAT TTTAAATGA AAGCGTAAAA ATGGTAAAAT ATATCAAAAT 1440  
 TGATTGTGAT A 1451

5

(2) INFORMATION FOR SEQ ID NO: 678:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 668 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 678:

15

nGTATTGAAG CGGTTAAACA AACACCTAAT GCAACTGACG AAGAAAAGCA GGCTGCTGTT 60  
 AATCAAATCA ATCAACTTAA AGATCAAGCA ATTAATCAAA TTAATCAAAA CCAAACAAAT 120  
 GATCAGGTAG ACACAACTAC AAATCAAGCG GTAAATGCTA TAGATAATGT TGAAGCTGAA 180  
 GTAGTAATTA AAACAAAGGC AATTGCAGAT ATTGAAAAAG CTGTTAAAGA AAAGCAACAG 240  
 CAAATTGATA ATAGTCTTGA TTCAACAGAT AATGAGAAAG AAGTTGCTTC ACAAGCATT 300  
 GCTAAAGAAA AAGAAAAAGC ACTTGCAGCT ATTGACCAAG CTCAAACGAA TAGTCAGGTG 360  
 AATCAAGCAG CAACAAATGG TGTATCAGCG ATTAATAATTA TTCAACCTGA AACAAAAGTT 420  
 AAACCAGCTG cACGTGAAAA AATcaATCAA AAAGCGAATG AATTACGTGc TAAGATTAAT 480  
 CAGGATAAAG AAGCAACAGC AGAAGAAAGA CAAGTAGCAC TAGATAAAAT CAATGAATTT 540  
 GTAAATCAAG CCATGACAGA TATTACGAAT AATAGAACAA ATCAACAAGT TGATGATACA 600  
 ACagTCagCG CTgATAGctT GCTTTAGTGA CGCCTGACCA TATTGTTAGA GCgCTGCTAG 660  
 AGATGCGT 668

25

30

35

(2) INFORMATION FOR SEQ ID NO: 679:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1906 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 679:

GATCCAAATT TAAAAGGAAA AATAGCCTTt AACGAATTtA CGAAACAAAt TGAATGTTTA 60  
 GGGAAAGTGC CATGGAATAC TAATTTTAAG ACACGTCAAT GGCAAGACGG TGATGATAGC 120  
 AGTTTAAGAA GTTATATCGA AAAGATTTAT GACATACACC ATTCAGGTAA AACAAAAGAT 180

50

55

EP 0 786 519 A2

	ATCGTGGGAT GGACATAAAC GCCTTGAAAA GTTATTTATC AAATACTTAG GTGTTGAAGA	300
	CcTGaAGTGA ATAGAACAAC TACCAAAAAA GCATTGACTG CTGGAATCGC TAGAGTAATG	360
5	GAGCCTGGAT GTAAATTTGA CTATATGCTT ACACTTTATG GTCCTCAAGG TGTAGGTAAA	420
	TCTGCTTTGC TAAAAAAATT AGGTGGTGCA TGGTTTCTG ACAGTTTAGT TTCTGTTACA	480
	GGTAAAGAAG CTTATGAGGC CTTACAAGGC GTTTGGCTAA TGGAAATGGC AGAACTTGCA	540
10	GCTACAAGAA AAGCTGAAGT TGAAGCTATT AAGCATTTCA TATCTAAACA AGTTGACCGA	600
	TTTCGTGTTG CTTATGGGCA TTATATTGAA GATTTTCCAA GGCAATGTAT TTTCATTGGT	660
	ACAACTAATA AAGTTGATTT CTTAAGAGAT GAACTGGTG GAAGACGTTT TTGGCCAATG	720
15	ACTGTAAATC CAGAGAGAGT TGAAGTGAAC TGGTCTAAAC TAACCAAAGA TGAGATTGAC	780
	CAAATTTGGG CAGAAGCTAA ACACTATTAT GAACAAGGAG AAGATTTATT CCTTAACCTT	840
20	GAACTAGAAG AAGAAATGCG TTCAATACAA AGCAAACATA CTGAGGAATC TCCATATACA	900
	GGCATTATTG ATGAATATCT TAACACACCm ATTCctAGCa ATTGGGATGA CTTAACTATC	960
	TTTGAACGAA GACGATTTTA TCaAGGTGAT GTTGATATGT TACCacAGG AAATGTaGAT	1020
25	TACGTTAAAA GAAATAAAGT CTGTGCGCTT GAAGTGTTTG TTGAATGTTT TGGTAAAGAT	1080
	AAGGGAGATA GTAGAGGATC TATGGAAATT AGAAAGATTT CAAACATCTT AAGACAATTA	1140
	GACAATTGGT CTGTATATGA TGGTAATAAA AGTGGGAAAA TTCGATTTGG AAAAGATTAT	1200
30	GGTGTACAGA TAGCTTATGT AAGAGATGAA AGTTTAGAGG ATTTAATATA AGAAATATTG	1260
	AATAAATATG CATTTTAGAG TGTTGTATCA GATGTTGCAT CATTTTGTGA GTGATGCAAC	1320
	ACGGGAGTGT AAAAAGTAAT CGTAGGTGTT GTATCATTTT TGGTGATGCA ACATTGATGC	1380
35	AACAAATGAT ACAACACCTC TTTCTTTTCT AGCTGTAGGG TTCAACCCTG TTTGTTTCCA	1440
	ATGTTGCATC AAATTCAC TAAGTTTAA AAAGTAGTGT TAGGGAGTAA AGGGGTATAG	1500
40	GGGTAACCCT CTAACAGCTA TTTTAAAAAG TTTGGCAAGA ATTGATACAA CATCGGAACA	1560
	CAAAATATAA TTTTGTATAC AAGGTGAATA AATGAAAGAA TCGACATTAG AAAAATATTT	1620
	AGTGAAAGAG ATAACAAAGC TAAACGTTT ATGTTTAAAA TGGGTCGCAC CTGGAACAAG	1680
45	AGGTGTGCCA GATAGAATTA TTATTATGCC AGAAGGAAAA ACATATTTTG TAGAAATGAA	1740
	GCAAGAAAAA GGAAAGTTGC ATCCTTTACA ArAATATGTG CATAGACAAT TTGAAAATAG	1800
	AGaTCATAAA GTaTATGTGT TATGGAATAA AGAACAAGTA AAaACTTTTA TCAGAwTGGT	1860
50	AGTGAACATT TGGCGATTGA CTTTCAAACC ACATAGCTnT CCAAAG	1906

(2) INFORMATION FOR SEQ ID NO: 680:

55

(A) LENGTH: 948 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 680:

10	AATTATTGA ATAAATTAGC AATTAAAGAG TTGATTGTG AGTTTAAGyA TTTGAGTGCA	60
	TTTGAAAAAG ATGTCATGTA TTTAATGTGT GAACAATATA AGCCGAGAGA AATTGCTCAA	120
	TTGATGCATG TAAAAGAGAA AGTGATTAT AATGCCATAC AACGATGTxA AAATAAAATA	180
15	AAACGTTATT TCAAAATGAT TTGAAAAGCG CCTTAGGACG TGAATTGAAT TATAACGTGT	240
	TACTTACTGA TGGTTTGACA TTTGTTATAA ATTTTATGTA TAGTATACTG GTATTATAAT	300
	GAATAAAGGT GAATTATTGT GAGAAAAATA CCTTTAAATT GTGAAGCTTG TGGCAATAGA	360
20	AATTATAATG TTCCTAAGCA AGAAGGCTCG GCAACAAGAT TAACCTTAAA GAAATATTGT	420
	CCAAAATGTA ACGCGCACAC AATTCATAAA GAATCGAAAT AAATACATTC GAAATAATAC	480
25	TTTGATAATA TGTTCAAAGG ATTTGGAGGT TGAGCAGATG GCTAAAAAAG AAAGTTTCTT	540
	TAAAGGCGTT AAGTCTGAAA TGGAAAAAAC AAGTTGGCCG ACGAAAGAAG AGCTATTTAA	600
	ATATACTGTA ATTGTAGTTT CTA CTGTTAT ATTCTTCTTA GTCTTTTCT ATGCCTTAGA	660
30	TTTAGGAATT ACAGCATTGA AAAATTTATT ATTTGGTTAG AGGAGTGAAG ACATGTCTGA	720
	AGAAGTTGGC GCAAAGCtTG GTATGCAGTG CATACATATT CTGGATATGA AAATAAAGTT	780
	AAAAAGAATT TAGAAAAAAG AGTAGAATCT ATGaATATGA CTGAACAAAT CTTTAGAGTA	840
35	GTCATACCGG AAGGAAGAAG GAAACCTCCA GTAAaAAGnT GGCCAAGCCT AAAACCGCCT	900
	GTTAAAAAAA ACCATTCCCC TGGGnTAnGG TTTTAAGTGG GAATTTAA	948

(2) INFORMATION FOR SEQ ID NO: 681:

40

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 863 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 681:

50	ACAnATAATA ACAAAGCGCT TGCTAGTACC TCTTAAAAAG ATGATGCTAG CAAGCGCTTk	60
	TCTATACTAT ATATTATTTT TCTAAAATTT TAACACCCTC TTGAGTGCCT ACAATAACTT	120
	GATCTGCCAT ATCTAAGAAG TATCCTGTCT CAAACACACC TGTCAGATGA ATTAAATACT	180

55

TATCAGTTAT AAATGCGACA TCTTCGTTTA CACGACGTTT TACTTTTATA TCAGCGTATG 300  
 ATTCAATTTT ACGTAATATG TGATACCAGT TAAATTTATC CACCTCTACT GGTAACCTTAA 360  
 5 ACGTCTCACC TAAGTATTGA ACTATTTTCG TTTCATCGAC AACCACAACA AAACGCGATG 420  
 CCATTTTCATC TATAACTTTC TCTCTGAACA GCGCACCACC GCCACCTTTA ATTATATTTA 480  
 AAGATGGATC TACTTCATCA GCACCATCAA TTGCTAAGTC GATATGATCA ACATCATTGA 540  
 10 TTTCACATAT TTTAATACCT AATTCTTTTG CTAAAAATGC AATTTTATTA GAAGTGCATA 600  
 CACCTGTAAT ATTGTAACCA CGTTCCTTAA TTAGTTGCGC CATTGAGGT AAGAGTAATT 660  
 CCATTGTAAT TCCTGTACCA ATTCCCAGCG TCATGTCACC ATTGATTkGA CTTAAAACAT 720  
 15 CATTTAATGT CATTAACTTG AGTGCTTTGA CATCTTTCAT GAAGGTAGCC TCCCATATTT 780  
 AAGTAATCTA TTCAATTCAT ATTTTACATG ACTCGTATAA ATTAACATAC CCTTATnGCT 840  
 20 AACCATTTGT GTTAAACATA TCG 863

## (2) INFORMATION FOR SEQ ID NO: 682:

(i) SEQUENCE CHARACTERISTICS:  
 25 (A) LENGTH: 480 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 682:

TTCAATTTTG TTGTCAGCA ATTTTGGTCC GAATTTCAAT TCATCAGATG ATAGCTCTAT 60  
 35 TAATTCGTGA GAATATTCTG CtACAACAAT TTCATAAATA TGGCCTTTTT CTTCCATTAT 120  
 TATTTTCATCa ATTATTTTCAT AATTCAATTG TTGTAATGTT TGTCTTAAAT TTTCAGTTTG 180  
 GATATTACTT TGTAAGTCA ACCTTGGATG TTGACTTAAC TTATCTTGCC CATCTTTTAA 240  
 40 AATTTTAGCA ATAAGTGGTC CGCCCATACC ACAAATTGTG aTATTATCGA TTACGTCCTC 300  
 AGGTGAATA AACTTAAGC CATCCCCTAA ACGTACATCA ATTCTATCTA CTAATTGGTT 360  
 TGCAGCTACA TTTTTCACAG CAGCTTGAAA AGGGCCTTGA ATAATTCTC CAGCAATAcC 420  
 45 GaTTCCGATA AATGGTTTTG AATTGCATAG ATTGGCAAAT AAGCATGATC TGAGCCAATA 480

## (2) INFORMATION FOR SEQ ID NO: 683:

(i) SEQUENCE CHARACTERISTICS:  
 50 (A) LENGTH: 689 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

55



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 683:

CTGCAAAAAA TATTGGTATA ATAAGAGGGA ACAGTGTGAA CAAGTTAATA ACTTGTGGAT 60  
 5 AACTGGAAAG TTGATAACAA TTTGGAGGAC CAAACGACAT GAAAATCACC ATTTTAGCTG 120  
 TAGGGAAACT AAAAGAGAAA TATTGGAAGC AAGCCATAGC AGAATATGAA AAACGTTTAG 180  
 GCCCATACAC CAAGATAGAC ATCATAGAAG TTCCAGACGA AAAAGCACCA GAAAATATGA 240  
 10 GTGACAAAGA AATTGAGCAA GTAAAAGAAa AAGAAGGCCA ACGAATACTA GCCAAAATCa 300  
 AACCACAATC CACAGTCATT ACATTAGAAA TACAAGGAAA GATGCTATCT TCCGAAGrTT 360  
 15 gGcCCAAGAA TTGAACCAAC GCATGACCCA AGGGCAAAGC GACTTTGTTT TCGTCATTGG 420  
 CGGATCAAAC GGCCTGCACA AGGACgTCTT ACAACGCaTa AcTACGCACT ATCATTcAGC 480  
 AAAATGACAT TCCCACATCA AATGATGCGG GTTGTGTTAA TTGAACAAGT GTACAGAGCA 540  
 20 TTTAAGATTA TGCGAGGAGA GCGGTATCAT AAGTAAACT AAAAAATTCT GTATGAGGAG 600  
 ATAATAATTT GGaGGGTGTT AAATGGkGGa CaTTAAATCC mCGTTCATTC mATATATAAG 660  
 ATATATCACG GTAATTGCGC ATATAACTT 689

25 (2) INFORMATION FOR SEQ ID NO: 684:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 858 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 684:

TTATTAAATT GGTATGTGTT CATTATACAT ATGaCAAATA TGaATGTAAA CCGATAATTT 60  
 AGATTTTTTG GAATAACCTG AAAATTCAAG TtaTAGCGTT GCTTATATTT TAAAAGGTGG 120  
 40 TGATAATGAG ACTTTTTGAA AAATAAAATT CAAAATACTT ATAGCATAAT CAATATGCAC 180  
 ATTAAATAAA TGTACTCTTT TAATGCGTTG ATAAGTGTAT TTGTAATTTA GAGAAGGGGT 240  
 GTTCACTATG CTGATGTGTT AAAAAATAAA ATAAAAAGGA CACCTCGATG CTATAAATAT 300  
 45 TAGCATCGAG ACGCCCAAGT AATGTCTATT AAATTGAATA TAGTCTCGGA CATGAATCAA 360  
 TGCCCTAGGC CCTGCAATGT TATATTGACA GTAGTTGACT GAATGAAAAT GACTTTGTAG 420  
 50 CTAGCTTTTT TCAATCCTTG TCGGTGCAAC ACATAGAGAA ATTGGATTCC TAATTTCTAC 480  
 AAACAATACA AGTTGCGGAA TAAGTCCCAA TATAGAAGGT GACAGTAAGC CAACTTACAA 540  
 TAATGTGCAA GTTGGTCGGG CCTCAATACA GAGATTTTCG AAAAGAAATT CTACATATTA 600

55

GCCACCTTGT TTATTTAAAT CGATAACACG GTTTGCGATT GTATTGATAA ATTCAAAGTC 720  
 ATATGAAGTA AAGATAATAG AACCTTTGAA TGATTTAAGT CCATCATTAA CAGCAGTAAT 780  
 5 ACTTTCTAAG TCTAAGGGTT GTGGGTCATC AAGTAAAGA CGTTGCCCTG ATAACATCAT 840  
 TTACTAGCAn CACGACTT 858

(2) INFORMATION FOR SEQ ID NO: 685:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1747 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 685:

20 TTAGTATTTT CAGGnTGGAA ATTGATGTTT GAAATTCACt GAAATGGGCC AGAACCTgAA 60  
 ATGaAACTTT ACCAACAATT TGGtCTTCAT CAATGAGGCC AAACGCACGG CTATCTTTAC 120  
 TTACTTCACG ATTATCTCCA AGCACTAAAT ATTTACCTTT TGAATGACA TTTGATTTAG 180  
 25 GATTGCGATT CGGTAAATCT TTAACCTGGA AAGTCCCAGT AATGTAATCA CCTTGTTTAT 240  
 GTTTTAAATT GTAGTTTAAA TATGGTTCAT CTGTGTTTTT ACCATTGACA TATAATGTAT 300  
 CATTTTTGTA TTCTACTTTA TCACCAGGAA CACCGATGAC ACGTTTAACA TAGTCATCAT 360  
 30 TTTTGTTTGC ATGGAAGACA ACTACATTAC CTTTTTCCAA ACCACCTGTT TTATATCCAA 420  
 CAATGTTTAC AGCTACTCGC TCGCCATCTT TCAAAGTTGG ATCCATTGAT TCACCTTTAA 480  
 35 TTGTATATGG CGTAACAATA AATTACCTA CTATAAATAA AATGACAAAA GCGACTGCAA 540  
 TTGAAATAAT CCATTCCAAT ATTTCTTTTT TCAATTTTGA CACCTCTTTT TAAGATTTGA 600  
 ACTGAACAGT CCATTTTGAA AAAGGATAGT ATCGTAAAc AACATTACCA ATAATAcCCT 660  
 40 TTTTATCGAT TAAACCAAAT TGTCTTGAAT CGTGCTTGtK ATTATCTTGA TCATTTAGCA 720  
 CAACAAAATT GTTTGGCGGA ATAATATCAC CATCTAATTC TTTAAAATTG CGCAAACtAA 780  
 AATCTTTAAT TTTTCTGTTC TTGGCATAAG ATGCGTCAAC CGGTCGGTCA TCACGGTATA 840  
 45 ATTGTCCCTG ACGAAACGCC ATTGATTGAC CAGGTTTGGC AATAATTGGA CTAGTATATA 900  
 TCTCGTTACC ACGCCTATAT GTAATGATAT CACCATTATT CAATTGATTA AATGTAACCT 960  
 50 TAATTTTATT TACAATAACA CGATCCCCTT TGTTAAGGGT TGGTGACATA TCATTATTCG 1020  
 GAATGACATG ACCAACTATT ACAAAGTTT GTACGAACAG TACAATGATA ATAGCAAGTA 1080  
 TCAATGAAAT CAAATATTTT ACAACTTTTT TCACGATGTC ACTCCTTTTT CGATCCCATA 1140

ATCCCTCTTA AAGGATGCGT TAAAATTGTA GTAATTCCT TACCTAAATA ACCTAAAATA 1260  
 ATTGTTGAAA CTAACCTTGA TGATGCCAAA ACAATGAAAT AATATTTAGG TCTAATATGA 1320  
 5 GATAGACTCG CTACAAAATT TATTAATGTA TTTGGCGTAA AAGGAAAACA AAGTAAAATA 1380  
 AACAAATGGA TTAATCCTTG GCGATCAATA AAATAATCA AGCGTTGAAC AGCAGTACGT 1440  
 TGTTTAATTC GCTGCATCCT CTCAGTGTTT ACCAATCGTT TACAGATCAA ATAGACTGTA 1500  
 10 AATGTTCCAG AAATTAATCC AAGCCAACTA ATCAATATAC CTAAAATAGG TCCATAAGCT 1560  
 TGAATGTTAA TTAATAATATA GAGTGCTAAA GGAAATACTG GAATTATAGC TCTAATATAT 1620  
 AACAAATATA ATCCAGGTAA ATAACCAAAC TGTCGAAATA TCTCAAACCA TTCTTCTACT 1680  
 15 TGATGAAACG ACAAATCATC AATCCCTTTC TTTGGTTGAA GATAATTATT CTTACATTAT 1740  
 AAAGTTA 1747

20 (2) INFORMATION FOR SEQ ID NO: 686:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 645 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 686:

CGTAAAGATT ACTATTTAGG AGGGTGAATA TGAAAAAGAA ATTAGGTATG TTAATTCTTG 60  
 TACCAGCCGT AACTTTATCA TTAGCCGCAT GTGGGAATGA TGATGGAAAA GATAaAGATG 120  
 35 GCAAGGTAAC AATTwAAcGa CaGTTaTCCm TTGCAATcAT TTgCAGAGCA AATTGGTGGA 180  
 AAACACGTGA AGGTATCATC AATCTATCCA GCAGGGACAG ATTTACATAG CTATGAACCA 240  
 ACACAAAAAG ATATATTAAG TGCAAGCAAG TCAGACTTGT TTATGTATAC AGGGGATAAT 300  
 40 TTAGATCCGG TTGCTAAGAA AGTTGCATCT ACTATCAAAG ATAAAGATAA AAAACTGTCT 360  
 TTAGAGGATA AATTAGATAA AGCAAAGCTT TTAAGTATC AACACGAGCA TGGTGAAGAG 420  
 CATGAACATG AGGGACATGA TCATGAGAAA GAAGAACATC ATCATCATCA TGGTGGATAT 480  
 45 GATCCACACG TATGGTTAGA TCCTAAAATT AACCAACTT TCGCTAAAGA AATTAAAGAT 540  
 GAATTAGTGA AAAAAGATCC AaAACATAAA GATGACTATG AGAaAACTA CnaAAATTAA 600  
 50 ACGACGATCT TAAGAAAATT GATAACGATA TGAAGCAAGT TACAA 645

(2) INFORMATION FOR SEQ ID NO: 687:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 956 base pairs  
 55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 687:

TTTGTTACTG	CTTCTAAAAT	AATATCCTTT	AATTGTTTAA	CATGTTGGAT	TGTCATATGA	60
GGTGATGGTA	CATTAAAAGG	ATTAAATTCA	TCTATTTGTG	CATATTGATT	TATGACATCT	120
TGATGCATTG	AAATAGGGTT	GATATCATT	GTTACTACTT	TATTAGATTG	GTCTTGTGAC	180
ATACTAATGG	TGCCACCAGT	ATGAATAACA	AGTAGATGTT	TCATATATTT	CCTCCTATAT	240
TTAATTTACC	TAATTATGAT	AAAATATTAT	TCATAAAACG	ACAAGGAAGG	GAAATGACGC	300
ATGAAAGCCA	TTAATATTGC	ATTAGATGGT	CCAGCTGCTG	CCGGAAAAAG	TACAATTGCG	360
AAACGTGTAG	CCAGCGAACT	ATCAATGATT	TATGTCGATA	CAGGAGCAAT	GTATCGTGCA	420
TTAACATACA	AATATTTAAA	ATTAAACAAA	ACTGAGGACT	TTGCAAACT	AGTTGACCAA	480
ACAACATTAG	ATTAACTTA	TAAAGCAGAT	AAAGGTCAAT	GTGTCATTTT	AGATAACGAA	540
GATGTAACAG	ACTTTTAAAG	AAATAATGAT	GTGACGCAAC	ATGTTTCATA	CGTTGCATCT	600
AAAGAGCCAG	TACGTTTCATT	CGCCGTTAAA	AAACAAAAAG	AGTTAGCTGC	AGAAAAAGGT	660
ATCGTAATGG	ATGGTCGCGA	TATCGGAACT	GTAGTGCTAC	CAGATGCAGA	TTTAAAAGTA	720
TATATGATTG	CATCAGTTGA	AGAGCGAGCA	GAAAGAAGAT	ATAAAGATAA	TCAATTAAGA	780
GGTATCGAAT	CAATTTTGA	AGATTTAAAA	CGTGATATTG	AAGCTCGTGA	TCAATATGAC	840
ATGAACCGTG	AAATATCACC	ATTAAGAAAA	GCAGATGATG	CAGTGACATT	AGATACGACm	900
GGCmAGTCCA	TTGAAGAAGT	TACTGACGAn	ATTTTAGCGA	TGGTGAGTnC	AATTnA	956

(2) INFORMATION FOR SEQ ID NO: 688:

(i) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 1166 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 688:

AAAAGGGAGG	AAAGAGAACA	GTAAATATG	AATACAAAAT	AATTTTGTTT	TCGGACAGCA	60
GGGGTATTAG	ACGCGATTGA	CAATGTC1GT	T1AATTAAAC	GTAATGTTTA	TTTAAGCGAT	120
GAATATTAGG	TGAAAAGTTT	TTGAATTGTA	ATGTAATTGA	GGTTTATTGA	TTAGACATTT	180
TATTGAATTG	CGTGTTATTA	TATAAATGTA	AAAATAAGAC	GACATGCGCG	AACATGTCGT	240

55

TAATGCGGAA TGGTTTTTTT ATTTCCGCT AATTGAAATA AAAATGACGT TTTAATATAT 360  
 TATGGGCTAG GTGGTTTGTA AGAAAGGGTT AGTTATTAAT GTTTTATGAA TTAAGGAAAT 420  
 5 TTAGAGTTTAA GGTTTAATCA ATTGTGATTT TGTTGATGAA GCGTTTAGTT AGAGTATTTT 480  
 CGCCACCACT AGTTACTCCT TCTCCCACTT TACCCGAGAC TGGAGAAGAg CTATCTGAAG 540  
 AATAAATAGA TACTTTTTTG CCATTTTGTA GTAAACCAAG ACCTTTTAAC TkCTCGGTTA 600  
 10 GAGAATTCCA TGTATTTTGA GCATCTAGCT TTTTGTTAAA GTCAyCGTAG ACATTTTCCT 660  
 TAGTTAAATC AATTTGTTTT AATCCTTTAA AGTCTATAGA TTGTGTTAGA TGGCCTCCAT 720  
 CATCATTTTc AGGAGCAGAA ACGTGCTAG AATATCCATT GCTTAATAAA TAAGTAACGT 780  
 15 TGATTGTTTC GTACTCGTTA CTTAAAATAA TATCAGAATC ATGTAAGaAT CTTTAACTTT 840  
 TTTCCATAAT TGACCATCTG TCATTTTTTC TTCTGCTTTA GCCGTTTTAA CAACTTTATT 900  
 20 TGTATCTAAT CCTAAGTATG AAGAATGTAA GCCTGTTCCCT AATGTTGTTA ATACTAAAGC 960  
 ACTTGCTACT aATGTTTTAC CTAAAAAtT TGTATTCATT TTTATTGCTC CTTtTTTTAT 1020  
 ATTGTAAACG TTTACAATGA AAATATAATA ATAATTTTTT AAAAGAACAA TTAATAAAT 1080  
 25 ATCAAAAATG TATTAATCTAT CTATTAATA AAAAATAGAA TAATTTTTTA ACATAGTTTT 1140  
 GTTGTTTTGA ATTATAAAAA CTAAAG 1166

(2) INFORMATION FOR SEQ ID NO: 689:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1083 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 689:

40 AGCACTGAAG GATGGCTAGT TGTCATAGCA TATGCGGTCA TTGGTATAGT TATAACGAGC 60  
 GTTCTTATA TGTTGTCTAT TAAATTTTTT AACAAACAAG AACTAKAATT GTCGATTTTG 120  
 GTGATGATAG TATTTGAAAT AATATTTAGA GCAGGTGATA AATCTTTACG ATTGTCATCT 180  
 45 GTTCTTTTkG GTGTGGAATG AAATGTGGGG GATAAGTATA GGTGACATAT CTATATTGAT 240  
 TTATTTGTTT TGAGGTGGTT ATGTTGTGTG GGAATTATTT CCTTTTAGAT AGCGGGGATT 300  
 AGAGGATATA TGTTATTTAT AAGTATCATT TGATGATTGT ATAGGCTAAC GATTCCTCG 360  
 50 GAAATATTTA AAAACCTCGA TCATGTAGCA TAACTGAAGT TTGTCACAAA AGTATAATGT 420  
 GAAGTTCGAC ACTTTTGgAT TCAGTTCAAA TACTTTGACC GAGGTAAATA CTATTTATTC 480

TGATACTGAG ATAATCATTa CATGGTCGTG ACCTTTAAAT AAAAGGCTGA CAATATAAGA 600  
 CATAACGAGT ATACCTAGTG AATATGAAAT ATACTTCGCG TTTGTCAGTT CATTATGGAA 660  
 5 ATAAGGCGTG ATTAACCATA ATCCAATATA GAATATTAAA AACTGATAT ACATCATATT 720  
 AATTTCAAAC AAGTCATTa GTTTATTGTT ATTACTAAAA ACAATTGCAG CATTAAATCAC 780  
 ACCTAAAGCG ATATTGATTA ATAGATGCGT ATACGATAAA CGGAAACCGA TAGATGTTAA 840  
 10 TTTATGATTA ATATAATTTT CAGTAATGAT CCAATATACA CCGAAAAGAC TAATTAAAAT 900  
 CATAAATTGG AATATATAAA TGTAATAAA ATGATCAATG CTAAATGATG ACGAAGCTAA 960  
 15 ACCAACCAGT ACCTCGCCAr AGWTATaATT GTTAGTAACG AAAAACGTCT ACTAAATGCA 1020  
 TCATATTAAC AGGTnTAATA CAAGTATTTTc TGAAATGGAA TAAGnCTGTC GCTGCATGAT 1080  
 ACG 1083

20 (2) INFORMATION FOR SEQ ID NO: 690:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 627 base pairs  
 (B) TYPE: nucleic acid  
 25 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 690:

TTAATCATCT GGATGTATTT AGTATTnAGA ATAATAAnAA AACGATCATG TTGTATtTGA 60  
 GTAGCGATTG GTTTGCGGAA TTAGGCTTTA CTTTCTTTAA TTACCACTAT ACAGCAAAGT 120  
 35 TGATTAAATC ATCCTATAAT TTGAAATGTC TACTATTAAA ATTGACATAT CGATACCTTG 180  
 ATAATCAGCC TCTTAATGAC GCTGATaYTA GAAAATTACA GGATATTATT AAAATCATTG 240  
 CAAAAGAAGC AAGTATGGAT AAAAAGATTG CACAAAATCA ATATCGATAT GCGTATTATG 300  
 40 GTGATTTGCG TGATGAGCTC GAATATATTT ATCAAAATGT AAATCAACGA TTGACATTAA 360  
 AAAGTGTCGC TGATAAATTA TTTGTCTCAA AGTCAAATTT GTCATCACAA TTCCACTTAC 420  
 TTATGGGCAT GGGTTTTAAA AAATATATTG ATACTTTGAA AATnGGTAAA TCGATTGAAA 480  
 45 TTCTACTTAC TACTGATAGT ACTATTAGCA ACATAAGTGA nCATTTAGGT TTTAGTAGTA 540  
 GCTCCACTTA CTCTAAAATG TTTAAAAGTT ATATGGATAT CACACCGAAT GAATATCGTA 600  
 ATTTATCAAA ATATAATAAn TGTTTAC 627  
 50

(2) INFORMATION FOR SEQ ID NO: 691:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 641 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 691:

	TTGAACGTAA TGCTAGCAAA TGACTTTGTG CCATAAAATA TTCTTCCCAT TTGATTCTTT	60
	CCAAGATGTT CACCTTCCAT ACTTAAATTT TAGTAACATT TTCTAATAAT ATAAGATTAA	120
10	TCACAAAAAA TAAATTTTGC AATTAAAAATA ATCCATTATG TCGTGAAATA AGATTTTCAGT	180
	TTATCAAAAG TTTTACTTCC AAAACCTTTT ACTTTTTTCA AATCGTCAAT TTCTTGAAAT	240
15	GCACCTTGTT GGTGCGATA TTCAACAATT GCATTAGCTT TAGCTTGCCC AACTCCAGGA	300
	ACAGACATCA ATTCTGATAC AGATGCCGTA TTTAAATTTA CTTTAGTATT ATTTGTGTTT	360
	CCaTTTTTTT CGTGCACT GTTTACTTCA ATTTGTGGTT CAACATTCTT TTGTCCTTTA	420
20	TGAGGTATGA AAATCATTTT TTGATCTGTT AATTTTTCAG ACAAATTAAT TTGACTTACA	480
	TCTGCATCCT CCAATAATTG TGCTTTATCA AGTAAATCAA CTACTCTATC CTTAGATGTC	540
	ATTTTATAAA CATTAGGATG TTTAACAGCA CCTTTTACAT CGACATATAC AGGACCCTTA	600
25	TTTTTGGAAAT TATCTCCATC TTGACCTGG ACATCTTCTA C	641

(2) INFORMATION FOR SEQ ID NO: 692:

30

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 631 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 692:

	TATTAAAGnA CTTGATGAAC CAAATCATAA AAAGChATAT ATGTTATTTG CAGCTGGCAT	60
40	TGTGTTnGCA ACTATTTTAC TTATTTTCGGC ACATTTATAC AGCAGAAAGA GAGGTAACCA	120
	AGTTTGAGAA TCATAAGTA TTTAACCAATT TTAGTGATAA GCGTCGTTAT CTTAACCAGC	180
	TGTCAATCTT CCAGTTCTCA AGAATCAACT AAATCCGGCG AATTCAGAAT CGTACCAACA	240
45	ACTGTTGCAT TGACAATGAC ATTGGACAAA TTGGATTTAC CAATTGTCGG CAAACCCACG	300
	TCATATAAGA CATTGCCTAA TCGTTATAAA GATGTACCGG AAATTGGTCA ACCAATGGAG	360
	CCGAATGTTG AAGCTGTTAA AAAGTTAAAA CCAACACATG TTTTGAGTGT GTCAACGATT	420
50	AAAGATGAAA TGCAACCATT TTACAAACAA TTAAATATGA AAGGCTACTT TTATGATTTT	480
	GATAGTTTAA AAGGGATGCA AAAGTCGATT ACACAATTAG GTGaTCAATT TAATCGTAAA	540

55

GCAGCTAAAC AAAAGAAACA TCCCAAAGTA T

631

(2) INFORMATION FOR SEQ ID NO: 693:

5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1111 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 693:

15	AATTTAACTA TGTTTTCCAC ATTGTTTCAT GTCACGAAAA GGACAACGCG CGACTATAAG	60
	TATCAACTAT TTCCACAAGT TTTATTGGTG TTTTATTAT TCATCGATAC GCTTCATTTT	120
	CATCTCTCCA ACACAAAAAA GAAGCTAAGC AACTTATGTT GCCTAACTCC TCTATACTAT	180
20	CCATATTTTA CTATTATCCA TATTTCAATG AATTATCTAA TGTGGGCTTC TATTTTTTCA	240
	ATATTTCTAC CGTCAATGAC GTCACCTCATG CGATTTGTTT GTAATTTTTT ATTAAGTTCA	300
	AACGTATAAT AGCCGCCATC TTTCATTATC ACTTTTATCT TACTATCTTT AGGAACTTT	360
25	TTATACAGAT CAAAATTTTG AATTAAATAC TGTCTCAATT TAAAGTCGAG TTCTTTAAGT	420
	GAAATCTCTT CTTTATAAAT GTAGTGACT CTACCGTACG TAGCAATACC GTCACCTTCA	480
	TCTCTCTTGA TTTGAAATCT TGGTGCGTTT ATATAATCAT AATAAGCGTC TTGATTTTTC	540
30	TTAGTGACAC CACCATATGA AAACACTGTG CCATTACGGT TTTCCGCTTC TTTAACAACA	600
	AATATGTCTA ATCCCGGATT TTTACGTGCT TTAAATCTTT CAATATCTTT ACCAAATATC	660
35	TGTACTCTTG TGAATTTTCT ATTTTATCA AAGATAAGGT AATGCTTGCC ACCTTTGCTA	720
	TAACGATAAC CAGTAACATT TTTAAGTTCC TTAAGTTGCG CACTATAGTA ATCTCTTAAG	780
	TCAAAGATAT CTTTGTGAC ATTTTCATAT TTTGCTTTAT GTTCACTCGC ATTTACAGTT	840
40	TGATGCAATG ACGTTATTGT TCCTGTTGCT AAAATACCTA ATGCTAAACT TGCTTTGCA	900
	ATTGCTGTCA TTTTCATAGT TGTATGCTCC ATTCGTAATT ATTAGATTG TTCGCTTACG	960
	TCTATTGAAT CATAAGCTT TATTATAGTT AGCGTATTTG ACCTTTCACA TTAAACCATG	1020
45	TTTAATAATC ATTGAATCAT TATTAAGTAA ATTAAGGATC TATAATGTTT GTTAAATAAA	1080
	CTGAnCCCGT TGTGCTTCAC ACCCGnTnGA T	1111

(2) INFORMATION FOR SEQ ID NO: 694:

50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 426 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

55



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 694:

5 TTATGGATGG ATTAAGAGGT CGTGTGAAA AAATCAACGA TAACTCTGTT ATTGTTGACT 60  
 TAACAATTAT GGAAATTTT AATGACCTTG ATTTACCGGA AAAAAGTGT ATCAATCATA 120  
 AACGATATAA GATTGTTGAA TAAGAAGGTA AGTTATAATG AATAAAATCT CGAAGGCTTT 180  
 10 AACTTGGTTT ATTATAAGTT TCATTATATT TCATCTCATA TTATTTATTA TGTGGGGCGA 240  
 ACACCAAGAA TACTGGTATT TATATACAGG TATAATGCTA ATTGCTGGTA TCAGTTATGT 300  
 15 ATTTTATCAA AGAGATATTG AATCTAAGCG GTTGCTTACA TCAATTGGTG TTGGTATTAT 360  
 TACGGCAATT ATTTTAATTA TGCTTCAACT TTTATKCTCA CTTATAAATT CTAATTTAAG 420  
 TTATAG 426

## 20 (2) INFORMATION FOR SEQ ID NO: 695:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 737 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 695:

30 AAAAGGGGnG TAAGGTTTAG CTCAAGTACG AGAAGTCTTT GGTGATGAAG CAATTGATGA 60  
 AAATGGTGAG ATGAATCGTC GTTATATGGG TGATCTAGTG TTTAATCATC CAGAAAAACG 120  
 CTTAGAATTA AATGCTATCA TACATCCTAT CGTGCGAGAT ATTATGGAAG AAGAAAAGCA 180  
 35 AGAATATTTA AAACAAGGAT ATAATGTAAT CATGGATATT CCATTATTAT TTGAAAATGA 240  
 ATTGAAAAT GCAGTAGACG AaGTGTGGGT TGTATACACT TCTGAAAGTA TACAAATGGA 300  
 40 TCGTTTAATG CAACGTAATA ATTTGTCAAT AGAAGATGCG AAAGCACGTG TCTATAGCCA 360  
 AATTTCTATT GATAAAAAA GCCGAATGGC CGATCATGTT ATCGATAATT TAGGGGATAA 420  
 ACTTGAATTA AAACAAAACC TTGAGAGATT GTTAGAAGAA GAAGTTATA TTGAAAaGCC 480  
 45 GAATTACGGA GAAGAAGATT AATATTACAC TATAATAAG TCATTACTTT ACGTACGCGT 540  
 TGATGTATGT AAGTAATGAC TATTTTTTAT AAAAAAGATA AATAAATCAA CGGAAAACGC 600  
 TTTCAAATTT CATATAATAT GCTATACTAA TTCCATAAAG TATAACACA AAAGATCAAG 660  
 50 GGGTGCTTTT AATGTCAACG AATATTGCAA TTAATGGTAT GGGTAGAATT GGAAGAATGG 720  
 TATTACGTAT TGCATTA 737

55

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 432 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 696:

10

AACGATAAGA	ACAAAAGATA	TATTACAAAG	CGTTTATTTA	AAACGTTATT	TATTACGCGC	60
GATGATGGCA	GGATTTATTA	TCGGGATTAT	TACGGTCTTC	GTATTATCAG	TTAAAGCAAC	120
ACACGAACCA	GATTTACCGC	CAGgCATTGT	GAATATGGCC	AGTGCCATTA	CATTTCAGCTT	180
15	TGCGTTAGTA	CTCATTTTAT	TTACAAACTC	CGAACTACTA	ACCAGTAACT	240
TACTGTAGGC	CTGTATTmTA	AAGTAATTAA	ACCAACTAGA	GTATTGaAAA	TATTTTTTATT	300
20	ATGCTTTGCA	GGAAATATTT	TAGGTGCTGC	TATTTyATTT	AGTTTCATGC	360
TGTAATGACG	CCAGATAwGt	TAAAyCAGTT	ATCAGCAGTT	ATAGAGCATA	AAACGTTGTC	420
TACTGGTTTT	GT					432

25

## (2) INFORMATION FOR SEQ ID NO: 697:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 782 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 697:

35

CTTTTATACG	AAAGTTAAaA	TCAAATATAA	CATTAATGTT	TGATGGGGAT	TTTGCGGGTA	60
GTGAaGCAAC	ACTTAAAACA	GGTCAAAATT	TGTTACAGCA	AGGGCTAAAT	GTATTTGTTA	120
40	TACAATTGCC	ATCAGGCATG	GATCCGGATG	AATACATTGG	TAAGTATGGC	180
TTACTGCTTT	TGTAAAAAAT	GACAAAAAGT	CATTTGCACA	TTATAAAGTG	AGTATATTAA	240
AAGATGAAAT	TGCACATAAT	GACCTTTCAT	ATGAACGTTA	TTTGAAAGAA	CTAAGTCATG	300
45	ATATTTGCGT	TATGAAATCA	TCGATTTTGC	AACAAAAGGC	TTTAAATGAT	360
TTTTCAATGT	TAGTCCTGAG	CAATTAGCTA	ACGAAATACA	ATTCAATCAA	GCACCAGCCA	420
ATTATTATCC	AGAAGATGAG	TATGGCGGTT	ACATTGAACC	TGAGCCAATT	GGTATGGCAC	480
50	AATTTGACAA	TTTGAGCCGT	CAAGAAAAAG	CGGACnAGCA	TTTTTAAAC	540
AGATAAAGAT	ACATTTTTAA	ATTATTATGA	AAGTGTGAT	AAGGATAACT	TCACAAATCA	600

55

TATCAGTGAT GCTGTGCAGT ATGTTAATTC AAATGAGTTG AGAGAAACAC TAATTAGCTT 720  
 AGAACAATAT AATTTGAATG ACGAACCATA TGAAATGAA ATTGATGATT ATGTCAATGT 780  
 5 TA 782

## (2) INFORMATION FOR SEQ ID NO: 698:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 714 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 698:

AAATCATTGG CAAAATATAC GAATTGCATT TAAATCGTCA ATTTACATAT ATTTTTCGCT 60  
 20 AGTAATCAAT CGTTATCATT GTTATTTATC GTTACATTAT TTCGAGTATC AGTATGTATT 120  
 TCGGGCTTCG TTTGATAACG ACATTTCTTT GTGACATCGC TTCATCAGTG TAACAACAAA 180  
 TACAATGATT TCGTGATGTT AGTTACCCAT TTTATGTGTT GCATAAAATA TGTGTGTATA 240  
 25 AAACATTTTA AATCATTTTA TATAACAAT CTATATATTT TTGGCATTTC CAAAATATCA 300  
 CTTGTTATAT TAAAAACCGA CAAGACATTT TATCTTATCG GTTGAAATTT GTTATTGTTA 360  
 TTTGTAATGT TTTTAGGTTT CTTTTTAATA TAATATATTT CAGTGAAAAT ACATGATTGA 420  
 30 TTGTGATTTT ACTGAAACAT GGTAAATTGC GTTGTTGATG AATAACTTTA GCATAAATAT 480  
 AGGAAGTTAT TTTGTACATC GCCATATATA GAAACGAAAT TATAATGACA GCTAGTACGT 540  
 AACTTGTTAA AAATATATGA TGGTTATTAA TACCTATCAT ATTTAGTAAC GTATATACAA 600  
 35 TGTACTAGA AATTAATGTG TGAATCAGTG CTA CTGTTAT TGGTATtGCG AACAGAAAG 660  
 TCATTTGATT TCGTGTTATC TTTGCTATTC TTCCATTATC TAAACCAAGT TTTT 714

## (2) INFORMATION FOR SEQ ID NO: 699:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 699:

ACTGAGAGCA ATAATTTAGT AACTTCTACT CAAGGAATTA TTAAAGAAGC ATTGCATAAA 60  
 TTGGGATTTG nntTTAATTA AAGAACCTTT AAGAATGTTA CAAGTGCGTA TCCCTGTACG 120

TGGACCAACA AAAGGGGGCG TCGTTTCCA CCCAGATGTT GATGAAGAAG AAGTAAAAGC 240  
 ATTATCAATG TGGATGACTT TGAAATGTGG CATTGTGAAA CTTACCATAC GGTGGTGGTT 300  
 5 AAGGGTGGTG ATCGTTTGTG GATCCACGTC AAAnGAGCAT TCCATGAAGT TGGAACGTTT 360  
 A 361

(2) INFORMATION FOR SEQ ID NO: 700:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 943 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 700:

20 GTCCTTAATT GGTATCCACC TTAGCACC CGTACACTTC GAATTAACCC CGCATTTCTT 60  
 AAAGGACCTA CAAGCTGTTC TAAATATAAA TCACTCAAAT TATTTTCTTC AGCAATTGAC 120  
 TTTAATGATA TACATCCTTG CCCCTCTTTT TTAGCAAGAG AAATCATCAA TGTAAGTCCA 180  
 25 TATCTCCCTT TAGTAGAAAT TTTCATTGTA TAACCTCACT TAATTCGAAT ATTGATATTC 240  
 CCATTTTAGC ATTTTTTGAG TTAAGATAGT ATAAGAAAGG TGTGACAAAT GTGAGTACAG 300  
 AACCATTAGC ATCGAGAATG CGCCCAAAA ATATAGATGA AATCATTTC CAACAACATT 360  
 30 TAGTTGGACC AAGAGGCATT ATCAGAAGAA TGGTTGATAC AAAAAAATTA ACTTCAATGA 420  
 TTTTTTATGG TCCACCTGGT ATAGGCAAAA CAAGTATTGC CAAAGCAATT TCGGGCAGTA 480  
 CGCAATATAA ATTCAGACAA TTGAATGCTG TAACTAACAC TAAAAAAGAT ATGCAACTTG 540  
 35 TTGTTGAAGA AGCTAAAATG TCTGGTCAAG TTATCTTGTT ATTAGATGAA ATACATCGAC 600  
 TAGATAAAGC TAAACAAGAC TTTTATTAC CTCATTTAGA AAATGGCAAA ATCGTCTTGA 660  
 TCGGTGCTAC AACTTCAAAT CCTTATCATG CTATCAATCC AGCGATTCGT TCAAGAGCGC 720  
 40 AAATTTTCGA GTTATATCCT TTAAATGACG AaGATGTGCG CCAAGCGTTA ACTCGTGCAA 780  
 TAGAAGATGA TGAGAATGGT TTGAAAmCAT ATCaACCCAA AATTGATGAA GATGCCATGA 840  
 45 CCTACTTTTC TACACAAAGC CAAGGTGATG TTCGTAGTGC GTTAAATGCA TTGGAATTAG 900  
 CTGTATTAAAG CGCAGATAAT GACAAAGACG GTTATCGACA TGT 943

(2) INFORMATION FOR SEQ ID NO: 701:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 445 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double

55

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 701:

5 CATTAAGTGA AGTTGTTGAT ACACCCATGC ATCAAGTCAA TTGTTCTGTT GATTTAGATA 60  
 CAGAAAGCTT ATTAGGCTTT AAAACAATTA AAACAAATGC GGAAGGTCAA CAAGAAATTG 120  
 TCTTTGTAGA TGGTCCAGTT ATTAAAGCTA TGAAAGAGGG GCATATTTTA TATATTGATG 180  
 10 AAATAAATAT GGCTAAACCT GAAACATTGC CTGTATTAAA TGGGGTCTTA GATTATCGTC 240  
 GTCAAATTAC GAATCCATAC ACTGGTGAAG TAATCAAAGC TGTACCAGGA TTTAACGTTA 300  
 TAGCAGCGAT AAATGAAGGT TATGTTGGtA CTTTGCCAAT GAATGAAGCA CTAAAAAaT 360  
 15 CGCyTTGTTG TtATTCACGT kGATTATaTT GATGGGGaCA TTTAAAAAAT GTGAnTAAGG 420  
 AGCAAGGTTT ATTACAAGAT GTTAA 445

## (2) INFORMATION FOR SEQ ID NO: 702:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 752 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 25 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 702:

30 TGCAAGATAA ACAAATTTGT CATTGTTTTG GTTGTA AAAA AGGTGGCAAT GTTTTTCAAT 60  
 TTA CTCAAGA AATTAAAGAC ATATCATTG TTGAAGCGGT TAAAGAATTA GGTGATAGas 120  
 TTAATGTTGC TGTAGATATT GAGGCAACAC AATCTAACTC AAATGTTCAA ATTGCTTckG 180  
 35 AyGATTTACA AATGATTGAA ATGCATGAGT TAATACAAGA ATTTTATTAT TACGCTTTAA 240  
 CAAAGACAGT CGAAGGCGAA CAAGCATTAA CGTACTTACA AGAACGTGGT TTTACAGATG 300  
 CGCTTATTAA AGAGCGAGGC ATTGGCTTTG CACCCGATAG CTCACATTTT TGTCATGATT 360  
 40 TTCTTCAAAA AAAGGGTTAC GATATTGAAT TAGCATATGA AGCCGGATTA TTATCACGTA 420  
 ACGAAGAAAA TTTCAGTTAT TACGATAGAT TTCGAAATCG TATTATGTTT CCTTTGAAAA 480  
 ATGCGCAAGG AAGAATTGTT GGATATTCAG GTCGAACATA TACCGGTCAA GAACCAAAT 540  
 45 ACTTAAATAG TCCTGAAACA CCTATCTTTC AAAAAAGAAA GTTGTATATAC AACTTAGATA 600  
 AAGCGCGTAA ATCAATTAGA AAATTAGATG AAATCGTaTT ACTAGAAGGT TTTATGGATG 660  
 TTATAAAATC TGATACTGCT GGCTTGAAAA ACGTTGTTGC AACAATGGGT ACACAGTTGT 720  
 50 CAGATGAACA TATTACTTTT ATACGAAAGT TA 752

55

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 830 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 703:

10

nCGTTTTTGT nAAACnGCTT ATATGTATAA CTTTTCCCAA TTTATCATAA GTTTTGATAG	60
AAGGAGTTGG ATCGCTAAAG AAAATATCTC TAAAGATATC ATCTTCAATG ATAGGAATAT	120
TATGTCtTTC GcTATAAGTA ATTATATTTT TCTTTTGCTC ATTCGTTAAA GAACGACCTG	180
TCGGGTtTATT AAACCTAGGT TCTATATAAA TCGCTTTATT TTAAAAATTA ATAAATCTAT	240
CAATGATGGT ATCAATTTCA TTAATTTGAT TATAAGGAAC ATCAATATGT CTAAAATTCA	300
ATTGCTCAAA AACATTTGTA GAGTGAATAT ATGATGGTGT ATTCGAAATT ATTATGGCAT	360
CTTGACCTAA AAACCCAATA GATAAAAGTT GAATGGCATG TAAAGCGCCT GAAGTGATCA	420
TTACATTTTC TCTACCTACA TTTATACCTT GCTTTGACAT TCGTTCAACG ATAATATCTC	480
TTAACTTGAT ATAACCATAG CCATTATTAT AACCAAAAGA TAAGTCTTCA ATATGACTGG	540
CTGTATTAGA CATGGCTTTT TTCAATTGAA TATGTGGCAT TAACGATATA CCAATTAC	600
CTTTACTTAT ATGTATATAC GAATCATCTG TCTCAATTTT ATTAATTAAT TGCACCGTAT	660
ACTGACTTCT TTGTTGAGAG GACCATAACA TCATTCAGA CCACTTATTT GTAATATGTG	720
CTTCATTCAA ATAGTCATTA ACATATGkTC CACTACCTAC TTTAGTATAG ATAAATCCTT	780
CAGCTTCTAA TAACTCAATA CTTTTAATAA TCGTTACTCT ATTTACGTTG	830

35

## (2) INFORMATION FOR SEQ ID NO: 704:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 659 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 704:

45

AGTGGTGTTG GAAAAGCTAT TATGAAATTA TTACGTGAAC AACAAGTTTA ATAAAAAAG	60
AGGGGTCAAA TATGAAAGGA TTAATTATTA TTGGCAGTGC ACAAGTGAAT TCACATACAA	120
GTGCACTAGC AAGATACTTA ACTGAGCATT TTAAACACA TGATATTGAA GCGGAAATAT	180
TCGATTTAGC AGAAAAACCG TTAAATCAAT TAGATTTTTT AGGAACAACA CCGTCTATTG	240

50

55

TATTAGGAAC GCCAACTAT CATGGTTCAT ATTCTGGAAT ATTGAAAAAT GCATTAGATC 360  
 ATCTAAATAT GGATTATTTT AAAATGAAAC CTGTAGGCTT AATAGGAAAT AGTGGTGGTA 420  
 5 TTGTTAGTTC AGAGCCATTG TCACATTTAA GAGTAATCGT CAGAAGTTTA CTAGGCATTG 480  
 CTGTACCAAC TCAAATAGCA ACACATGATT CTGATTTTGC TAAAAATGAA GATGGTTCAT 540  
 ATTACTTAAA TGATAGTGAA TTCCAATTAC GAGCAAGATT ATTTGTCGAT CAAATTGTAT 600  
 10 CTTTTGTGAA TAATAGTCCA TATGAACATT TAAATAATA TTAAnAAATA TGTAAATnT 659

## (2) INFORMATION FOR SEQ ID NO: 705:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 400 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 705:

ncaccatata gtaactggca ccaactacat taccgtcttt taaaaagatt tttttatagt 60  
 25 TATTATCAAC ACTATTAAAT ATTTCAATAC CTTTAATTTC TGCATTTTCT ACAATTTGAC 120  
 CAGCACTATA CAAGTCACAC CCAGAAACTT TTAATGACGT AAATGTTGTT GATCCCTTGT 180  
 ATCCGTTTCGT TTCTTTATTT GTTAAATGAT CAGCTAATAC TTTACCTTGT TCATATAGTG 240  
 30 GTGCAACGAG TCCATAAACT TTGCCGTTAT GTCTGCACAT TCACCAACTG CATATACATT 300  
 GCTATCACTT GTTTGCATCA CATCATTGAC AACAAATACCA CGATTACATC TAGACCTGAT 360  
 CTTGGCACTC CTGGGAAGGC GGAACCACTG CATACTACTA 400

35

## (2) INFORMATION FOR SEQ ID NO: 706:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 380 base pairs  
 (B) TYPE: nucleic acid  
 40 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 706:

TTAAAAATAC AGCTACAGGT AATTTTAATG ATTTTTCATC AATATCAAAT TTGGGATTAT 60  
 GGTGTGGCGC TGTAATACCT TTAACGAGTC AGAAAGAATG CACCTGGTCG TACTTTCAAA 120  
 50 TAATGTGAAA AATCTTCTCC AATCATCATT AAATCTGATT CATTAAAGCG TACATGTAAG 180  
 TCATTTGTTG CTTCTTTAAT AACTTGGATA TGCTTTCTCG TTTATTATGG ACAGGCAAAT 240

55

GCTTAATCCA TTTGTCCAT ACATGATTCT GTATATCTGA AATCGAAAGT TCTnACTGTA 360  
 CCTTTACAAA ATGCCTTGnn 380

5 (2) INFORMATION FOR SEQ ID NO: 707:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 (B) TYPE: nucleic acid  
 10 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 707:

CAnAAAGGAT CAAAGTGATC GGTATAGTGA TAATTAGCTC CAAAGAAAGA ATATTCTAAA 60  
 TCAAATCCAT ACCAAGCAGA AAGTATTAGC GAATATCAAT TTAAATGGTA CCGATTCAAA 120  
 20 TAAAGAAACA CGACATATAG AATTTTACT TGATGATTTT AGTGAATCAT ATGAACCAGG 180  
 AGATTGTATA GTAGCATTAC CGCAAAACGA CCCTGAATTG GTTGAAAAAC TAATATCCAT 240  
 GTTAGGTTGG GATCCGCAAT CTCCGGTGCC AATTAATGAT CATGGTGATA CAGTTCCTAT 300  
 25 TGTGAAGCA CTAACATCAC ATTTTGAATT TACTAAATTA ACATTGCCAT TATTGAAAAA 360  
 TGCAGATATC TATTTTGACA ATGAAGAATT ATCTGAnCGT 400

30 (2) INFORMATION FOR SEQ ID NO: 708:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 447 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 708:

40 TAATGCTGGT ACGGGTCATG CAGCATTATG TGAGTTGAAC TACACAGTTT TACAACCTGA 60  
 TGGTTCTATC GACATCGAAA AAGCGAAAGT GATTAACGAA GAGTTTGAGA TTTCAAAACA 120  
 ATTCTGGGGT CACTTAGTGA AAAGCGGTAG CATCGAGAAC CCAAGAGAAT TTATCAATCC 180  
 45 ATTACCACAC ATCAGTTATG TTAGAGGTAA AAACAATGTT AAATTCTTAA AAGATCGTTA 240  
 CGAAGCGATG AAAGCTTTCC CTATGTTTGA TAATATCGAA TATACTGAAG ACATCGAAGT 300  
 AATGAAAAAA TGGATTCCAT TGATGATGAA AGGCCGTGAA GATAACCCCTG GTATCATGGc 360  
 50 GGCAAGTAAA ATTGACGArG GTmCAGATGt AAmCTyCGGT GAATTAACAC GTAAAATGGC 420  
 TAAAAGCATT GAAGCACATC CAAATGC 447

55



## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1448 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 709:

10	GTGTCIACCT GTTTTGTG TTCAATTGTT AACTTTTCTT TTTGAATAGT AGTATTCGAT	60
	AATTCITTTAT CGCTTATAAA ATCATCTAGT GGATGGCTTT CTTTGAAACT TTTATTTTCA	120
15	GTCATCAATC ATCTCTCCAA TAGTTGCTAT TTTACATCAA TCTTGATGAT GTTTTTGATG	180
	TAACITATTT AATTCCATTT CAATATCTAA ACGTTCATAA TCATCTTCGT TGAGACGCTT	240
	TAAATCAGCG ATTAATGTTT GTTTGACCTC ATCCAAAGTA ATTCGTGTTT GTTCTAACTT	300
20	TTGCTGTTCA TTAATTGATT TTTTGGGCAT TTTTGCTAGA CGTGATATG CATCAACCAA	360
	ATTTAAAGCA TTATCAATAT GAGAATAAAA AAAGCCTTCA ACTTTATAAA ATGATGCAGG	420
	TCTCTGTCTA ACTGTCGTAT AAATAGAACG TGAAATTTGG TATATATCAT TAATCTGCCT	480
25	AAAATCTTTA ATTGATCTTA TATTGACATA CGTTTTTAAA ATACCTCTAA GTTTTTGGTG	540
	TGTATGATTT AACTGATTTT GAATATAGCG ATAGTCTTTT CTAGTCAAAC CAATTTCTGT	600
	TAAATATTTG CGTGAAGTGA GTTTTTGTAT CGGTAGGTAT GTCATTAAAA AGCCAACAAT	660
30	ACCAATAGAC ATATCAATTA AAAAAGATAC ATCAAGTGCA ATCATCCCAA ATATGCTTGT	720
	TAAAAACGCT ACAGGAATTC CCACTAACAC CCCAAATATA TGAGAAATAT TATATCTCAC	780
	TGTCATCTTC CTTTATTTAG CATTTTATAT TGATCGAAAA TCCGATAATT TTTGATTTAG	840
35	TTCTAACTCT TCAAGTTGAT GGCTTGTTAC ATTTGATGCT GGTGAGGCAC CTTCAATTAC	900
	ACCTTGAATA AATCTCTCTA TATCTGCGTC ATCCCCCTGT GCATATATCT CTACATAGTC	960
40	ATCTACATTT TGAACAGTAC CGACAATGTT ATAGTTCATT GCAATGCGTT GTGTAAAAATA	1020
	TCTAAATCCG ACGCCTTGAA CGCGTCCGAA TACTTGTAATA TGTATATGtC TCaTTTTTAC	1080
	CACCTCATAA TGTTATTATA CGTAGTTTTA CTTAAAAAAA CTAATAATTA CTATAGTTAC	1140
45	TACTTTGTTT GTTCAAGTC GTCAAACCTG ATTTTCAGAG GATAAAGGTA TAAAAATAAG	1200
	TATAGAGTTT TTGAAGTATG GAAGGGGTCT TTAATAATGT GGACAGTTAC CAAAATTAGA	1260
	GCCGATTATG AGGGATGGTG GTTATTCAGT GACTGGCCAG AAAACATTGT TGAAAAATAT	1320
50	CAATATCAAG ATTTTGATGA CATGTTTAAG CACTATCAAC AATTGATTAA TCAATGTAAA	1380
	GTTCAAGTTCG ATAACATATGT CACAGGCAAA TATAATATTT ATGCATTTTA TAATAATTGT	1440

## (2) INFORMATION FOR SEQ ID NO: 710:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 519 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 710:

GCTGnAAAAT ATGTTAGAAG CAGATGCATG GGCAAAGCTG GGGTCCTTAT TTATTGTCAG 60  
 GTAAAGATGT CTTCAATTCA ACTATTGGAA TATATGGTAT GGGAGATATT GGTAAGCTT 120  
 TTGCAAGAAG GTTGCAAGGG TTTAATACTA ATATTCTTTA TCATAATCGA TCAAGACATA 180  
 AAGATGCAGA GCGGACTTT AATGCAACAT ATGTTTCTTT TGAAACGTTG TTAGCAGAAA 240  
 GTGATTTTAT CATCTGTACA GCGCCACTTA CAAAAGAAAC ACATCATAAA TTTAATGCTG 300  
 AAGCATTGTA ACAAATGAAA AATGATGCAA TTTTATTAA TATCGGTAGA GGACAAATTG 360  
 TAGATGAAAC AGCATTAAATC GATGCACTAG ACAATAAAGA AATTTTAGCA TGTGGTTTAG 420  
 ATGTATTAGC AAATGaACCG ATTGATCaTA CACATCCATT aATGGGaCGT GaTAATGtTC 480  
 TGaTTaCACC aCACATTGGG TAGGCGCATT CAGTTAACh 519

## (2) INFORMATION FOR SEQ ID NO: 711:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1349 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 711:

CTATTACAGG TGGCGAACCA ATGTTTCTA AAAAGTCTAT TAGAAATGTT GTTAAACCTC 60  
 TATTAAAGTA TGCACATCAT CGAGGTATAT ATACACAAAT GAATTCAAAC CTAACATTGC 120  
 CTCAAGATCG TTATTTAGAT ATTGCTGAAT ATATCGATGT TATGCATATC TCACACAACT 180  
 GGGGAACAAC TGATGAATTC GCAAATGTTG GCTTTGGCGC AATGAAGAAG CAACCACCGT 240  
 TAAAAGCTAA GTTAAAATTA TATGAACAAA TGATTTGAA TGCACGTACA TTATCAGAAC 300  
 AAGGAATGTT TGTATCTGCG GAAACAATGC TCAATCAAAG TACGCTACCA CATTTACGAA 360  
 AAATACATCA AGAAGTCGTT CATGATATGA AATGTAGCAG ACACGAGATT CACCCTATGT 420  
 ATCCAGCTGA CTTTGCAAGT CAATTAAATG TGTAACTCT AGCGGAAATG AAAAAGACAA 480

TGTITCCATG CTTAAAGGAT GATGAAGATC AAAAGTTACT ATCACGTTTA AGAAATGCTA 600  
 AAAATGTAAC GACTAGAAAT GACCCGGATG GCCGTAGTCG TTAAATGTC AATGTATTTA 660  
 5 CAGGTAATGT AATCGTAACT GATTTCGGAG ATGAAACAGG TACAATTTCG AATATACAAA 720  
 AAGATAAATT AACAGATGTA TTGATAAAT GGTTATCCTC TGATCTTGCT AAATCATTA 780  
 ATTGTCATTG TTCCGAGTTT AGTTGTTTAG GGCCAAATGT TCTTGTTAAA AATATGTACT 840  
 10 ATCCGAATAT GGATTTTAAA GATAATGAGC GTCATATGCA CAAACAACCA CAAATTATAC 900  
 AATTTTAAAA ACTCTTAATT ATGCGGAGAA GCACTTTATC GATAAGTAGT CTCCGCATAT 960  
 TTTAATGCTA TTATAAAATA AAAACAATT AATTGCTGGC AGTACTCTAC TTAAATAATA 1020  
 15 AAGGGCATT T AATAGGACTA ATAGTCTATA ATAAAGGGG TAAATTTTAA CTAAAGCAT 1080  
 AAACGTGCAT AATCAAAAAG ACAGaTTGTA GGTGGAATAT TCGAaCATAA CAGTTCAATT 1140  
 20 CATCCTTAAC AATCTGTCTT TATATTTTTA GTCTCTAATA TGTTGCACTT GAGCTAAATA 1200  
 TTCAATTGTT gTTTACTTtC AaTGCGaCgt GCTTtCtTc GTtCAaCaCG TgwGGTGCTG 1260  
 TATCATAAAA CCATTTTTCa ACATCATCTT CTGGATATAC ACCAGGTACA TGTTTAGGTT 1320  
 25 GCCTTCATCA TCTAACGCAa CAAATGTAA 1349

(2) INFORMATION FOR SEQ ID NO: 712:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 377 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 712:

GAAATTAAAA AAGCAATTGG ACAAGATGCA ACAGTGTGAT TGTGATGA ATTTGATAAA 60  
 40 AAATTATACA CTTACGGCGA TAACTGGGGT CGTGGTGGAG AAGTATTATA TCAAGCATTT 120  
 GGTTTGAAAA TGCAACCAGA ACAACAAAAG TTAAGTCCA AAAnCAGGTn GGCCTGAATG 180  
 GAACCAGGAG GAAnTTGAAA ATATGCTGGG GATTACATTG TGAGTACAAG TGAAGGTAAA 240  
 45 CCTACACCAG GATACGAATC AACAAACATG TGGAAGAATT TGAAAGCTAC TAAAGAAGGA 300  
 CATATTGTTA AAGTTGATGC TGGTACATAC TGGTACAACG TCCTTATACA TTAGATTTCa 360  
 TCGGTAAAGA TTAAAAA 377

(2) INFORMATION FOR SEQ ID NO: 713:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 528 base pairs

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 713:

	AATGAAnCGA AAAATGACTA TAAAAAGTTT AAAGTGT TTT CACTTATTTT AACACTTGTC	60
10	ATTGTCATTT TAGCAATTAT AAGATTGTGTT CATAAAATGA TGTAATTAGA GTGAGACATT	120
	GTTTTATGTC TCAGGATCCA GTTATTCACT ATATCTACAA TATTTACGAT TATATAAATA	180
	ACCCGAGATT TTAGTATGAT TCATTtCACT AAAATCTCGG GTTCTATTT GATAATTTTT	240
15	AATGGGATAT GGCATGTATA CGTTCCTGCC TTTTATCTCA TTTCCAATGA TTAATCTGGA	300
	TATTGTTCTA AAAATGCTTT CGCTTCTTTA TTAAGTGT TAAAATCAAT ACCTTGTTGC	360
	ATCGCTGCAA AGACACATCC ACAATAACAC TGCCTAAAGA TATTATAGTC ATTACACATT	420
20	TCTATGGaAC GCTCATAACC TTTACTTTTC yTAAATCAC TTGGcAAATA GTTCACaTCG	480
	TATATTTTTT GGACATCCAT ACCAAGTTCA TTGATTAATT GTGCGTTC	528

25

(2) INFORMATION FOR SEQ ID NO: 714:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 731 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 714:

35	TTTAATGGCA TGAAGTCACA GTATATAAAG AATCTGGTGT GACAATCAGT ATGACTAAGT	60
	ATATAAAAAA AGAGCAGGTG cGAAATAATG GCGAAAGAGT CGAAATCAGC TAATGAAATT	120
	TCACCTGAGC AAATTAACCA ATGGATTAAA GAACACCAAG AAAATAAGAA TACAGATGCA	180
40	CAGGATAAGT TAGTTAAACA TTACCAAAAA CTAATTGAGT CATTTGGCATA TAAATATTCT	240
	AAAGGACAAT CACATCACGA AGATTTAGTT CAAGTTGGTA TGTTTGGTTT AATAGGTGCC	300
	ATAAATAGAT TCGATATGTC CTTTGAACGG AAGTTTGAAG CCTTTTGTAGT ACCTACTGTA	360
45	ATCGGTGAAA TCAAAAGATA TCTACGAGAT AAAACTTGGA GTGTACATGT TCCGAGACGT	420
	ATTAAAGAAA TTGGGCCAAG AATCAAAAAA GTGAGCGATG AACTAACCGC TGaATTaGaG	480
	cGTTcACyTT CTATcAGTGA AATAGCTGAT CGATTAGAAG TCTCAGAAGA AGAAGTGTTA	540
50	GAAGCAATGG AAATGGGACA AAGTTATAAT GCGTTAAGTG TTGATCATT CATTGAAGCT	600
	GATAAAGATG GTTCAACTGT TACGCTATTA GATATTATGG GGCAACAAGA TGACcTTAT	660

55

CGAGAAATCA T

731

## (2) INFORMATION FOR SEQ ID NO: 715:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 830 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 715:

TTTTAGrGGa AAGCGaATTA GtCaTATwCg CaGCAGATwG aATGATTGaa AAmGaAATTG 60  
 aTTCGATTCC AATTGTAAGA AAAAAAGATA ATCAAAAGTA TGAAGTAATT GGAAGAATTT 120  
 CCAAAACAAC AATAGCTAAG TTATTAGTAG CATTATATAA AGAATAGGTG AGAAGTAATG 180  
 GAAAAAATTA AAATTATCGT AGCTTCAGAT TCTATAGGTG AAACGGCAGA GTTAGTTGCT 240  
 AGGGCAGGTA TTTCACAATT CAATCCTAAG CAATGTAAAA ATGAATTATT AAGATATCCA 300  
 TATATTGAAT CTTTTGAAGA TGTTGATGAA GTGATTCAAG TTGCAAAAGA TACAAATGCT 360  
 ATCATGTGTTT ATACACTTAT TAAACCTGAA ATGAAGCAAT ATATGAGTGA GAAAGTAGCA 420  
 GAATTCCAAT TGAAGTCTGT CGATATCATG GGGCCATTAA TGGATTTATT ATCTGCTTCG 480  
 GTTGAAGAAA AACCTTATAA TGAGCCAGGT ATCGTTCATA GATTAGATGA TGCATATTTT 540  
 AAGAAAATTG ATGCGATAGA GTTTGCAGTT AAATATGATG ATGGTAAAGA TCCTAAAGGA 600  
 TTACCTAAAG CTGATATTGT TTTACTTGGT ATTTCGAGAA CTTCAAAGAC ACCATTATCT 660  
 CAGTATTTAG CGCATAAGAG TTACAAAGTT ATGAATGTAC CGATTGTACC AGAAaGTGAC 720  
 ACCGCCAGAT GGCTTATATg GATATTAATC CAAAGAAATG TATCGCACTT AAAATAAGTG 780  
 AAGAnAAATT AAATCGCATT AGAnAAGAGC GACTAAAACA ATTAnGACTA 830

40

## (2) INFORMATION FOR SEQ ID NO: 716:

45

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1159 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 716:

nGAAGAACAA GTTTAnCATC TAAATGCCCG TTTAAACTA CAACTACTTG ACGATGTTAA 60  
 ATCAGTGTTT AAnTCTCAAA TGACGCAAAA TAGTGATTTT AATGAAGAAA AGAAAGTGTC 120

55

	AGAACGTATA AAAAAATACT TTAATAAGCA ACTCACTGAG CAAATTGCAC CAATCGTTCA	240
	ACAATTAGCA GATTTACATG TCATTATTAA TCCTCAGTTT AACTTTGAAT CAGCTAATAT	300
5	AGAGCAACCA TTATTGCACA TCGATTTCAA CGATATGCTA AATGCATTGC CTAAACAATT	360
	AACAAAACGT AAAATTTTGA ATCCAAATGG GCAAAGAGAT ATACATGAAT CAATTTGTCA	420
	AAGTACGTTA GGATTATTAC AACCACAAAT GGGATTATTG AGGCAACAGC TTGAATTATA	480
10	TGTAAAGCAA ATGGCTGTAG AAGCTGAATC GCAATTTGAA AGTTTTGAAG CTAATATTCA	540
	AACGCAAATA AACGATTTAT TAGCATTTGA TTTAGATACA ACACTTATCA ATCAATTGAA	600
	AGATAAACAT CAACAACCTGA AAACATTTTT ATATTAAGAA AGAAGGAACG TTTTAAATGC	660
15	CTAATAAAAT ATTACTTGTA GATGGTATGG CGCTATTATT TAGACATTTT TATGCTACAA	720
	GTCTTCATAA ACAATTTATG TACAATTCAC AAGGTGAACC TACAAATGGA ATACAAGGAT	780
20	TTGTGCGTCA TATCTTTTCG GCAATACATG AAATACGCCC TACACATGTA GCTGTATGTT	840
	GGGATATGGG ACAATCAACT TTTAGAAATG ATATGTTTGA TGGTTATAAG CAAAACGTT	900
	CTGCACCACC AGAAGAATTG ATACCACAAT TTGATTATGT TAAAGAAATT TCAGAGCAAT	960
25	TTGGCTTTGT AAATATTGGC GTTAAAACT ATGAAGCGGA TGATGTTATA GGTACATTAG	1020
	CACAACAATA TTCAACTGAT AACGATGTCT ATATTATTAC GGGCGACAAA GATTTACTGC	1080
	AATGTATTAA TGACAATGTT GAAGTTGGCT AATTAAAAAA GGTTTAACAT TTATAATAGA	1140
30	TATnCATTAC ATCGTTTTn	1159

(2) INFORMATION FOR SEQ ID NO: 717:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 574 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 717:

	GATTTAATCA ATGCAGTTGC AGAGCAAGCT GATTTAACTA AAAAGAAGC TGGTTCAGCA	60
45	GTAGATGCTG TATTCGAATC AATCCAAAAC TCACTTGCTA AAGGTGAAAA AGTACAATTA	120
	ATTGGTTTCG GTAACCTTGA GGTACGTGAA CGTGCTGCAC GTAAAGGTCG TAACCCTCAA	180
	ACTGGTAAAG AAATTGATAT CCCAGCAAGT AAAGTTCCAG CATTCAAAGC TGGTAAAGCA	240
50	TTAAAAGATG CTGTAAAATA ATTTTACTTA AAAAGCCCTG AATAAGGGCT TTTTATTTTG	300
	CTTTTAATAC TTACAACCTG TACATAAATT GTAATGTTCT TCTAAGTTT TAATCTTTGG	360

AACTACAAAA TATACATATG aATATTGaGa TTAATTGTTA GCGTTGaATT TACTTAAAAG 480  
 GTAACCATGT CTACTATAGT ATTTTACGTT ATTTAAAAAG ATGAATAATG TAAATGAAGT 540  
 5 AAAGGTTATT ATGAGAATTA CAAAAGCTAC ATAA 574

(2) INFORMATION FOR SEQ ID NO: 718:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 493 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 718:

AAGaTATTGA AAmTGMaATG GaGCACGCCC TTATTTGATA GAAGTAAAAG ACATTTAATT 60  
 20 CTTACCGATG CAGGTCAAAT TTTTATAGAG AAAAGTAAAG AAATTGTtGC ACTGTATGAT 120  
 TATTTACCAT CTGAAATGGA ACGCTTGAAT GGACTGGAAA CAGGACATAT AAACATGGGC 180  
 ATGTCGGCAG TCATGAATAT GAAGATTCTT ATCAATATTC TTGGTGCAAT CCATCAACAA 240  
 25 TATCCAAATG TTACATATAA TTTGATAGAA AATGGCGGTA AAACAATTGA ACAGCAAATT 300  
 ATCAATGATG AAGTAGATAT AGGCGTGACC ACTTTGCCAG TCGATCATCA TATTTTCGAT 360  
 TATACTACCC TAGATAAGGA AGATTGCGA CTTATCGTGA GCAGAGAGCA TCGACTCGCA 420  
 30 AAATATGAAA CTGTTAAACT CGAAGATTTA GCAGGTGAAG ACTTCATTTT ATTTAATAAA 480  
 GACTTTTACT TGA 493

(2) INFORMATION FOR SEQ ID NO: 719:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1257 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 719:

GCATTGCCAA AACAATACGT TTATATTATT GAGGAACTAC TGTATAAAAG TAATGAATAT 60  
 45 CAAAATAAAA AATCATATTA CGAAACACTT GTTAACCAAG TAATTGAACT TAAACAGGCA 120  
 GATGATTTAA TTATTGGACT CGCTTATTCC GTACAACGCT TAGTCGTCGA TCATTTACAC 180  
 50 GTTGTGGTG ATATTTATGA TCGTGGACCA CAACCAGATA AAATTATGGA TACACTGATT 240  
 AATTATCATT CCCTAGATAT TCAATGGGGT AATCATGATG TGCTTTGGGT TGGAGCCTAT 300

GATATTATCG AAGACGCTTA TGGCATTAAAT TTAAGACCAC TGCTTACTTT AGCTGAAAAA 420  
 TACTATGACG CAGATAATCC TGCTTTTAAAG CCTAAAAAAA GACCTGACAA ACACGAACGT 480  
 5 TTAAGTCAAC GTGAAGAAAG TCAAATTACT AAAATTCATC AAGCTATTGC GATGATTCAA 540  
 TTCAAGTTAG AAATACCAAT TATTAAACGT CGTCCAAAT TCGAAATGGA AGAACGTCTT 600  
 GTGCTTGAAA AGGTTAATTA TGATACAAAT GAAATTACAG TTTATGGTAA TACATACCCA 660  
 10 TTGAAAGACA CATGTTTCCA AACTGTCAAT CGTGATAATC CAGCAGAATT ACTACCTGAA 720  
 GAAGAAGAAG TCATGAATAA ACTATTATTG TCATTCCAAC AATCTGAAAA ATTACGTCGT 780  
 CATATGTCTT TCTTGATGCG TAAAGGCTCT CTTTACTTAC CATATAATGG CAATTTACTC 840  
 15 ATTCATGGTT GTATTCCAGT TGATGAAAAT GGTGAGATGG AATCATTGGA AATTGATGGT 900  
 CATACTTACA GCGGCCAAGA ATTATTAGAT GTGTTTGAGT ATCATGTCCG TAAATCATT 960  
 GATGAAAAAG AAAATACTGA TGACTTATCG ACGGATTTAG TTTGGTATTT ATGGACTGGG 1020  
 20 AAATATTCGT CACTATTTGG TAAACGTGCC ATGACTACGT TTGAGCGATA CTTTATTGCA 1080  
 GATAAAGCTT CTCATAAAGA AGAAAAGAAT CCGTACTATC aTCTTCGTGA AGATGTGaAT 1140  
 25 ATGGkTCGTA AAATGCTCaG TGaTTTCGGA TTAAATCCAG ATGAAGGACG CATTATTAAT 1200  
 GGTCACACAC CAGTGAAGA AATCAATGGC GAGATCCTAT CAAGGCTGAT GGAAAGA 1257

(2) INFORMATION FOR SEQ ID NO: 720:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 720:

40 TGAGCACATG ACAGTGGCTG AAAACATTAA ATTTTTTAAA TCACTTTGTA AAAATCCAAT 60  
 TAACGATACA ACTATCAACG AATATTTACA GCAATTAAAC TTTGATGATA CGTCTGCCAA 120  
 AGTATCTACA TTGTCCGGTG GGAATAAACG TAAAATTAAT ATATTAGTAG GTTTACTAGG 180  
 45 TCAACCTCGA ATTCTCATTT TAGATGAACC GACAGTTGGT ATTGATTTAA AATCTAGACA 240  
 TGACATCCAC CAACTACTTA ACATCATGAA ATCTAAATGT TTAATTATAT TAACTACCCA 300  
 TCATTTAGAT GAAGTTGAGG CACTTGCAGA TGGTTATCAA GTTAATTGGG CCAGGTnCCC 360  
 50 TTTTnTTCAA CAGTTTTTGG GGGCCAACCA TGGGCTTATA 400

(2) INFORMATION FOR SEQ ID NO: 721:

55



(A) LENGTH: 570 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 721:

10	CTTATTATA TTATAATATA AATATTTTAT TTTTCAATAC TTAATTTTAA AAATCCCCTA	60
	TGAAAATGTT GATGAGCTAC ATCTTTAAAT CTTATTATAC ATTGTATAAA ATTATATTGC	120
	GAGGTAGTAA ATTGATATTA TACACTTTTG GTTGTGTATG CATCTCACCT ATCTTTTTTG	180
15	CTTTTTTCAA ATAAGAAAGT ATTAAAAATy AATGCCTATA CAGGAACAGA CATGGCTAAA	240
	CTTGATTTAA ATAGTCTTGA CGACGAGCAC GTAAAATTAT TAATAAATGA ATTAAAATAT	300
	CCAGAACTC ATATCGATGT AAATGAATTA AAAACAATAG TTGCTAGTCG AATAAATGAA	360
20	AGGCAAGAAA TAATAAGTTT TAAGTTAGGA ATAAAGTACT TATTAACAAT AAAAAAGAGGG	420
	AACATAGAAA AAGATAGGTT TTCAATTTC AATCATTTTCA AAGATACCTA TCACACCCTA	480
	GTTAGAATAG ATATTAACGG TGGTACTCAC GATAATCCAG ATGGAACAAT CGCTCCGAAA	540
25	AGTCATATTC ACATATATAA TGATAAGTAT	570

(2) INFORMATION FOR SEQ ID NO: 722:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 722:

40	ATTTTAGTCT TAGGTGTTGA TTGCATGATG AATGCAGAAG TTATACCTGC AGCCATATTA	60
	GCACCTTCAT TATTGTGTAT AATTGAATA TTATTGCTT TTAAAGTATG TCCAATTAAG	120
	TTTGAAGTCG TTGTTTTACC ATTTGTTCCA CTGATAAATA CAATATCATC AACTTGCTCT	180
45	GCTAATTTTC TTAATATATC TGTATCCACT TTTCTAGCGA TTTGTCCAGG TAAATCTGTT	240
	CCTCTTTTAC CTA CTGCTCT ACTTGCTTTA CGCGCCAATT TCGCTAGATG GATTGCCGTC	300
	CACTGTCTCA TGTGTTTCCT CctCAAATt CCACTCGCAT CATTATAACA TGACAAGGCA	360
50	ACTTCAAAAA AGTTTCTCAA TCACAAATnG ATACCAGTGT	400

(2) INFORMATION FOR SEQ ID NO: 723:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1113 base pairs

55

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 723:

	GATTATATGC TGTGGCAAA ACAATATACC CAGATGAATT TTTATTTTGA GTAATAGTAT	60
10	AACCATGCTT TGTCTTACTA ACTTTGACAG CTTCTAATGA ATTTGAAGCG TCTTTAAGCG	120
	TGGTATAATC TTCGCCGTAT ATCCCTTTTA AATTTACTCG ATACTTACCT TTAGGCAATG	180
	ATAATCTAAT TCTATCTGGA GCTTTAATGC GTATCGTTAC GGGTGTTACA ACGCGTCGAT	240
15	ATTTATAAGT GAGTTTATTT CTTTCTTG TG TATATTCATT CACTTTAACA TCATGAGCTT	300
	TATCCGGCGA AAGTAATTCT AAATCCATTT CAAAATACAA ATCTTTAAAT TGATTAGAAA	360
	CTGATTTTGG CAACTGTACA GTTAGACCAC CATTATTTTG TTAACTTGT AATAAATGTT	420
20	TTGTAGGAGA TTGCCAGGCT GCACTATTTA ATTTAATTGT TGAATCTGAT AGTAAATTTT	480
	TATTGGCTTT AAAATGTGTA TTAACATCTT TAATATTGTT AGAAACAATC CCTTGCAACA	540
	TTGCTTGTTT TTTATCTAAT GGAGATTTTA ATTCTTTATT GGAAAAGACC TTATTTGTAA	600
25	TATGTGCACT TGGATAATGG ATGGTATTTT TAGAATGAAT CCAACGAACT TTATTGTCTT	660
	TGTGTTTCTG CTTAATTTTA AATCCATATG GtAAGTTGTC ATCATGATtC ACTCTAATTC	720
30	GATCATTAAC ATTCCAAAGT GATAGTtAAT TTgACGATgC CAAGTAATCT ATAAGTGCTG	780
	TTTTTATCGA TTGGCATATT AATTTGGAGT GTCTTGTCAT AATATTTTAA AATGTCTCCA	840
	TTAAAAATAC TAGAATATAA TGAAATGCCA TTATAATGAT ATATAAATGG TGAATTTAAT	900
35	GCATAGTCTG ACATATAATC AATGCGATTA AATGAGCCTG TTGCATTTTG ATTTATCTTT	960
	TTTATAAGCT GGTttACATA GTTACTATGG TAATCATGTT GTttCAACGT TGATAATGAT	1020
	TGTTGATAAG GTTTGATTGC CATGTTTTTG TTGTTATCTA AAATGACGAT TTGTTGAATC	1080
40	ATaACGATTA ATACTAATAT TGCAACGGTT AAT	1113

(2) INFORMATION FOR SEQ ID NO: 724:

45

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 464 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 724:

55

TGCCGTTTAG CAAAATCTAG TATCGCTTGA TGATCAGATT CTGAAATTTT AGTGTGTACT	60
---	----

GATTGATTAA GTTTATATGC AAGTGCATGT TCTCGTCCAC CAGCACCAAT TACTAATACA 180  
 TTCATTTATG TTACGCCCTT TAAAAATTAG TGTTTAAAT GTCGAGTGCC TGTGACTACC 240  
 5 ATTGCAATAC CATGTTTATT AGCCATATCA ATTGAATCTT GATCTTTAAT CGAACCACCC 300  
 GGTTGGATAA TTGCCTTTAT ACCATGTTGT GCTGCAAGTT CAACTGTATC TCCCATAGGG 360  
 AAAAAATCCAT CAGATACTAA CGGCTACATG ATCATTGATT TCAATAGCTC TCTCTAACGC 420  
 10 AATTTTAGCA GCACCGACAC GATTCATTGG CCAGCAnCTA TACC 464

## (2) INFORMATION FOR SEQ ID NO: 725:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 1440 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 725:

GTTAAATnCA nTATTTnAAA TaTATTACCT TATTAGAAAA AGTCGTAATA TGAGGTGTAC 60  
 25 AAATGACGCA AATTTTAATA GTAGAAGATG AACAAACTT AGCAAGATT CTGAATTGG 120  
 AACTCACACA TGAAAATTAC AATGTGGACA CAGAGTATGA TGGACAAGAC GGTTTAGATA 180  
 AAGCGCTTAG CCATTACTAT GATTTAATCA TATTAGATT AATGTTGCCG TCAATTAATG 240  
 30 GCTTAGAAAT TTGTCGCAAA ATTAGACAAC AACAACTAC ACCTATCATT ATAATTACAG 300  
 CGAAAAGTGA TACGTATGAC AAAGTTGCTG GGCTTGATTA CGGTGCAGAC GATTATATAG 360  
 TTAAGCCGTT TGATATTGAA GAACTTTTAG CAAGAATTCG TGCAATTTTA CGTCGTCAGC 420  
 35 CACAAAAGGA TATTATCGAT GTCAACGGTA TTACAATTGA TAAGAACGCT TTTAAAGTGA 480  
 CGGTAAATGG CGCAGAAATT GAaTTAACAA AAACAGAGTA TGATTACTA TATCTTCTAG 540  
 40 CTGAAAATAA AAACCATGTT ATGCAACGGG AACAAATTTT AAATCATGTA TGGGGTTATA 600  
 ATAGTGAAGT AGAAACAAAT GTCGTAGATG TTTATATAAG rTATTTACGA AACAGTTAA 660  
 AACCATACGA TCGTGACAAA ATGATTGAAA CAGTTCGTGG CGTTGGGTAT GTGATACGAT 720  
 45 GACAAAACGT AAATTGCGCA ATAAGTGGAT TATTGTTACC ACGATGATTA CGTTTGTCAC 780  
 GATATTTTTG TTTGTTTTAA TTATTATTTT TTTCTTGAAA GATACACTGC ATAATAGTGA 840  
 GCTTGATGAT GCaGAACGAA GCTCaAGCGA TATTAATAat TTATTTCaTT CTAAGCCTGT 900  
 50 TAAaGATATA TCTGCaTTAG ACTTGAATGC aTCTTTAgGT AAtTTTCaAG AGATAATTAT 960  
 TTATGATGAG CATAATAATA AATTATTTGA GACATCGAAT GATAACACAG TGAGAGTTGA 1020

55

ATATTTAATT ATTAAAGAAC CAATTACAAC GCAAGATTTT AAAGGGTATA GCTTGTTAAT 1140  
 TCATTTACTA GAAAATTATG ATAACATCGT AAAATCATTG TATATCATTG CGCTGGCATT 1200  
 5 TGGAGTGATT GCAACAATTA TAACTGCCAC AATCAGTTAT GTATTTTCAA CACAAATTAC 1260  
 TAAACCGCTT GTCAGTTTAT CAAATAAAAT GATTGAGATT CGACGAGATG GTTTTCAAAA 1320  
 TAAATTGCAA TTAAATACAA ATTATGAAGA AATAGATAAT TTAGCAAATA CGTTTAATGA 1380  
 10 GATGATGAGC CAAATTGAAG AATCATTTAA TCAACAAAGA CAATTTGTTG AAGATGCGTC 1440

(2) INFORMATION FOR SEQ ID NO: 726:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 375 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 726:

TGATATTCCG ACTTGCAGAG ATTATTGCGC ATTAAGGATT AACGCAGTAT GAGTCCAAAG 60  
 25 ACGCGCGACA CCTGGAAAAA GGATTACCTA ATGCCTTATT TACAGTAACC TTGTATGATA 120  
 AAGATCGGTT AATTGGTATG GGTAGAGTGA TTGGCGATGG CGGAACTGTT TTTCAAATTG 180  
 TTGATATTGC AGTTTTGAAA AGTTACCAAG GTCAAGGTTA CGGCATCTAA TTATGGAGCA 240  
 30 TATTATGCAA TATATTAAAG GTGTGGCTGT TGAGAGTACA TACGTTATCT GATTGGCAGA 300  
 CTACCCAGCG GGnTAAATTA TATACCAAAT TTGGGTTTnA TACCTACCGA ACCAGAnTCC 360  
 AGGCGGTGAT GTATG 375

35

(2) INFORMATION FOR SEQ ID NO: 727:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 471 base pairs  
 40 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 727:

TAAAAAAGCTT AGAGATGAAA ATCATATTAT TTATCTGTTT TTGGGACAAT TTTTACTAA 60  
 AAATGAAGAT CCATGGCATC AAATACTTAA TGATTTAGAA GTTACAAATT CTGTTGATAA 120  
 50 TTTTTTAAGG TCAATAAGTA ATAAGGCCAA AGAAACAAAA AAAAGAGCTT TTATTATTAT 180  
 TGATGCGCTT AATGAAGGTG AAGGTAAAAG GTTATGGGGA AATTATTTTC AAAGCTTTAT 240

55

AGATGTTATA TTACCCAAAA ACGCAATACA AGATAACAAT ATTGTAGTAT TTCAGCATGA 360  
 AGGTTTTAGT AAGGAAGAAA ACTATAATCC AATTGTATCT TTTTGTGATT TTTATGGATT 420  
 5 AGAGCTACCT AAGTTACCTA TATTAAATCC AGAATTCAAC AATCCATTAT T 471

(2) INFORMATION FOR SEQ ID NO: 728:

(i) SEQUENCE CHARACTERISTICS:  
 10 (A) LENGTH: 1750 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 728:

TAGGATGTTT CACCCTgGCA ATTGcTCTAA GACATcTACa GCAGTAGAAT ayCCTGTTTT 60  
 20 CGTcTtTTTA ATAACAGGTA ATTGTAATGT CTCAAACAAT ACAACACCTA ATTGCTTAGG 120  
 AGAATTTTATA TTAAATCTT CACCAGCTGC ATCATGGATA TTTCGAATCA AGACGTCTAA 180  
 TTTTCTTGA ATTTCTTTTT CCATTTCTTC TAAATCATGA ACATCTGTAA ATATACCAAT 240  
 25 TTCTTCCATT TCACTTAAAA TCTTAGCTAG CGGTAGCTCT AAATCAGCCA AGAGTTCTAC 300  
 CTGATTGTAT TCTTCTAATT GTTTATCCAT ATTTGGTTTC GCAAAGTAAA TTGCATCAGT 360  
 AATAGAAGCA ACATATGGAT TTAAACATC ATCTTCAGGT ACCTTAAATT TCTTACCTTT 420  
 30 TCCATATATA CTCACATCGT CTTTCACAAA ACTTTGACCG TACAATGAAA CAACTGATTG 480  
 AACATCACTA ATCGTACGAG ATGGATCAAT TATATACTG GCCAACATAA TATCGAAAGA 540  
 AATATTTTGA ATATCAATCC CCAATCTATG TGATGCTACA TATGTTTTTT TAGCATCATA 600  
 35 TACGACTTTT TTCGAATTCG GATTTTCTAA CCATGAAACT AGTTCGACAT AATTATTTAT 660  
 GTCATCCGCA TTAATTACAA TATGTTTCTC ACCTGtAAAT aAAGAGAATT TTAAAtATT 720  
 40 ATTTcGCAAA TAGtTACCAC CGtCTAATTC GAAATGGATG GCCGCTTCTT TcAATGAAGT 780  
 AAAATCcAAT ATTATCAAAA GACGTTTCCa ATTTCAAATG TCTTtTCTAT TGCATCTTCA 840  
 ACGcTtGCTG ATTGaTCAAT GTCAGCCAAC AATTGTTTGA ATTCTAACTT CTAAACAAT 900  
 45 TCGATTTTTT CTTGTTGTTC ATCTtGakGA GTCATTAAAG TATCTTCaAG TTTTACTTCa 960  
 ATCGGACTAT CTACATTAAt CGTTGCTAAT TCTTTACTCA TTAATGCATC TTCTTTGCTA 1020  
 TTTTGAAGTT TTTCTTTTAA CTTTTTACCT GAAATTCAT CTAAATGTTT ATAGACACCT 1080  
 50 TCTACTGTGT CAAATTGGTT TAGCAATTTT ATTGCTGTTT TCTCTCCAAC ACCTGCAACA 1140  
 CCTGGTATAT TATCAGAAGT ATCTCCCAT TATCCTTTCA TATCAATAAT TTGATTAGGT 1200

55

CTTTTTTTAG TGTAATAAAT GGTTACATTA TCCGTTGCAA GTTGTGTTAA ATCTCGGTCT 1320  
 CCCGTAATAA TAATTGTCTG AAAGCCCGCT TTATCTGCTT CTTTACTTAA AGTTCCGATA 1380  
 5 ATATCATCTG CCTCATAGTT ATCTAATTCA TAACGTTTAA TATGATAAGC ATCTAATAAT 1440  
 TGGCGAATAT AAGGAAATTG CTCACCTAGT TCAGGCGGCG TTTTCTGGCG TCCACCTTTA 1500  
 TATTCACATAT ATTTTTCATG TCTGAAAGTC GTTTTACCTG CATCAAACGC TACTAAAAAA 1560  
 10 TGATTTGGCT TTTCTTCTTT TAAATCTTC TCTAGTAACA TTGCAAACC ATATACTGCA 1620  
 TTGGTATGAA TGCCTGCTTT GTTTGATAAC AAAGGTAATG CATAAAAGC TCTAAAACCT 1680  
 AAGCTATTAC CATCGATTAA TACTAATTTA TTCACAATTT TAACCTCCAG AACTAATTTA 1740  
 15 TATATnTGTG 1750

## (2) INFORMATION FOR SEQ ID NO: 729:

20 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 439 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 729:

AAATTAATTT TAACATCCTT TCAAATAGT TTTAACGGAT CcTCCcAAAA CGTAAACTCA 60  
 30 CATCAGTTAC TTGTAACATG CATTTTCTCC TTTTTTTCAT TCGATATTCT AACGGAAGAA 120  
 TTATATCATA TTATCGTCAC AGTTTCGACC TCATATAAGT TGTAATGATA GAATGACTCA 180  
 CACATGTTAT AATAATAAAG AATACAAGAA TCGAAGGAGA ATAAACATGG CATTAGACAA 240  
 35 AGATATAGTA GGGTCTATAG AATTCCTTGA AGTAGTAGGG TTACCAAGGT CAACTTACCT 300  
 TTAAAGGAC CAACGGTGAA ACGTAAAGTT AAACCAATCA GAATGAACGT GATGATTGAA 360  
 40 TTAGAAGTAG GGGAGnATAT AGTTCCTCAT TTATCCAACC GGTCAGGTGG ATTATTGCAC 420  
 TCAAATATGC nGnTATACG 439

## (2) INFORMATION FOR SEQ ID NO: 730:

45 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 481 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 730:

55

AGTTTAGGAT TGATGACAAA GCCATTACGA CCAACAAATC CTAAACCTGC ACGTTCTGCT 120  
 ACTGCCCTAT CTGATAATAC ACCCGTATCT ACCATAGATT TGATTTCAAC ATCTGGAAC 180  
 5 TTAGATTCAA TAAATGCAGC TAACATGTCT AATCGTTTAC GCATAATTGT ATGATAATCT 240  
 TGACCCACG ATGCTCTAGC AAATAAGCCT CTGCGATCAC CTCTAACACT CTTAGGTGCA 300  
 CCTTTCAGTT TGTTAGGATA ACCAACTGCA ATTGCTATGA TTGACCTTGC TGTTGGTAAG 360  
 10 GATAATTTAG GTCCTGTCG TAAAGCAATA TCAGATTCTT CAAaTCCTGA GGCATAACCA 420  
 TTTGCATGAT ATGCTTCTAG CTTTGTCTC AATTCATCAA AGGGnTCGGC AGTAGTAAAT 480  
 15 C 481

## (2) INFORMATION FOR SEQ ID NO: 731:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 828 base pairs  
 20 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 731:

GTGATGAAAA ATCTTTTCT ATTGATTGAG AAAATAGTTA TTTATATAGT TTTTAAAGAA 60  
 AAAAAGTGAT AATATTTAGC TAATCTAATG AAAATTGATT TATTGGACAA AAAATACATT 120  
 30 TTAAATGAGT AGAGGAGGCT GCAAWTGGGT TACAGAACTT TAAAAAGTAT TTTTCATGAA 180  
 CACAATGAAA GTAAAAAGAA GGAAGAGTAT ATTAAAAGAT TTAATTCTTT mGCTTCTTTC 240  
 AATACTAATA TTAATATCAT ACCTATGGAA AATGGAAAAA AAGTTAATGA TTTGGAATAT 300  
 35 CCTCTATTCT TTATGGTGAC TAAAAATCTA TCAAAAAAAC AAGAATTAAT ATCAATTAAT 360  
 AGTAGAAAAA TTGATAGAGC ACTTAATTCT TTACCATATG CAGCTAGAGA ACAATATTTT 420  
 40 AATGATTTAT TAATCGATGA ATTACAAAGT ACTAATGAAA TTGAAAATGT ATTTAGTACT 480  
 AAACAAGAGA TTGCACATGC GTTAAATAAC CAAGCATCAG AATTTCTTAA GTTCAGAGGC 540  
 CTCGTGGATC AATATAAAGA GATAGAACTT AATAAAAAAA TTAAAGTTGA TAATGTAAGA 600  
 45 GACATTAGAG CGATTTATGA TAAATTAGTT TCAAAATGAAA TTAACGAACA AGATAAGTTA 660  
 GATGGAGAGC TATTTCTGTA AAATTTTGTC GGTGTGCATG ATGGGTCAAC GAATAAATAT 720  
 ATACATGTTG GGTTACAACC TGAAACCAAA ATTGTTGAAT TTATAGGTGA AATGCTAACA 780  
 50 TTTTAAAAAT ATTTTGATGC GCCTCAGCCG TTCAAAATCA TGGCTAGT 828

## (2) INFORMATION FOR SEQ ID NO: 732:

55

(A) LENGTH: 1622 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 732:

10	CGTATGTTTC GAATTTTATG nATTGCATTG GATAATAGTA CCAGTTAAAA GAACTAAGAT	60
	TGTAATTATC CCCCCTTAAA ATTCAAGAAT ATATTTTCTT AATCTATCAG CAAGAGGTTA	120
	TATTTCAAAG TGTTTATTTT TAAAACAACA TCTAAAAGCC ATTTcATAAt AAwGTAAAT	180
15	CaTTAGaATG TATAAGATT CCaATTAATA AAAATAGATA AAATGCAATA AAGTTCCAAA	240
	TACTTATTAT TATTCATCAC AAATTACGTG ATGCCCTCTA CAACACTAAA TCAAGGATGA	300
	TATTTATTTT ATATACATTA GCATTCATTG TTTGCTTACC CACATTATCC TGTTTAGTCT	360
20	TTTTTATAAC TCTAGTCTTG AATGTCATAT TTACAAAGAA AAAAACACTA AAATATTTAA	420
	AACTAACACT ATTTATTAC TTTGTTCTTT TATAGCACCT TTTATGTTTA TTTTCTTGTA	480
	TTCAATAAAT AGTATGGAGA TGATAATTTA AAGTAGATAT ATTTGAAGTA TACTATAGAG	540
25	TAAATTATAT TGGGGAAATA TGTAACGAA CCGGTTGATT TTATAGTGGA TTAACACTTC	600
	ATAGAGAATA TAATCAAGAA CAAAAACAGT CAATAGGTGT GAATTTTTTT AAAGACGGAT	660
30	CTGTTGCTAC TAAAAAAGA AGTATACAAT TCAATAGTTA AACCTTAAAA CAAGAAATAT	720
	TATTCAAAT CAATGAATTT CCTATCTTAT TAGTTTTAAC AATATTTATT CTCATAGAGT	780
	TTGTCCAATT AAATATAGAT GATTCAAATA TTTATAAGCA TACAAAAGAG CAGTAAGACA	840
35	TTTTCTAATA GAAAATAACT TTAGTGCTCT TTGTACATTA CTCATCAACT ATTGTAAATT	900
	AAATAATAAA TAACTACCTA TTTTATTATT CAGCTAAAA ATGCTCTGCT AAATATTTTG	960
	CTGCGCCATC CTCTTCATTC GTGTATGCCG TTACATCTGA AGTTAATGCT TGGATTTTCA	1020
40	GGCGTGCATT TTTCATAGCA ACTGTATAAT GACCAAATTC AAACATTGCT CTATCATTGT	1080
	CGCTATCTCC AATAACTAAC GTTTCTTCTT GATGAATACC AAAATGTGCA ATCATTTCTT	1140
	TAATGCCTGT ACCTTTATCA GTTTGATAAG CCATTGTTTC CGCATTAAAT CTTGATGAAT	1200
45	TTGArACACT AATCTGTAGT TGCACATGAT TTTGCTTTAA TTCATCTCTA AATGCTGTTA	1260
	TTTTTTCTAA ATTAGAAGT AATAGATAAA TTTTGAATA TTCACCTTCA GGAAACTTAG	1320
50	TTACCCAATC TATCTTACCA GCAAGCGCAT CTTGTCTTGA AGACCATTCA CTATGACTTA	1380
	CGCCATTAAT AGGATCTTGA CTACGAATCA TATCTCGCAT CCATGTTTCA TCTTCTTTTA	1440
	AAGAACTCT ATTACCTTCA AAAGGAAATA CCTCATAATA AATTTGTTGG CGCTTAGCTA	1500

55



CTACTTCTCC AATTGTTCCA TTGAACTAA TGATGCCATT AACCGCAAAA TCTTGAGGTA 1620  
CA 1622

5 (2) INFORMATION FOR SEQ ID NO: 733:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 472 base pairs  
(B) TYPE: nucleic acid  
10 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 733:

AATGCCAAT AAATCCAAAT TCTTTTAACG AACAAGAGTG CCCAAGTTAT TGCTGATCGT 60  
TATGCCAGAT AnGCATATCA ATGATAATTA TGGTTTAGAA AGAATTCTA AGACAAATCA 120  
TGGATATAAT TATGTGTATT CCAATGATAA TTCAACTAGT AAGCAACATG TAAGTATTTT 180  
AAATCAAGGC ATAATAACGA AATAATAGAT GGAACAGTGT ATTCTAATTG GATATACTGT 240  
TTTTATTTTG CAATAATTTA ATTTAAAAAG GTGAATTCAA CTTATAAAAT GATGTAAATG 300  
25 TTATGTCAAA ATCAACCAAT CCGTAATGTA TTTTAAATG TTAATATAGT TCTGAAGAAG 360  
TATAAATGAG GTGTGAAAT GGCTAAAAAT AAGAAAACgA ACGCGATGCG TATGCTTGAT 420  
CGTGCAAAAA TTAAATACGA AGTTCATAGC TTTGAGGTAC nTTAAGAACA TT 472

30 (2) INFORMATION FOR SEQ ID NO: 734:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 519 base pairs  
(B) TYPE: nucleic acid  
35 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 734:

CATAGCCCCA AATTTTTTCT ATTATTTGTT CACGAGTAAA GATTTGCTTA GGACGTGCTG 60  
CAAGCATAAA TAATAATTGA AATTCCTTGT TCGGTAACGT CATCGTTTTA TTAGATACTT 120  
45 GGAGTTCCAA ATAGGATTGG TTTAGCGTTA AGTTGCCAAT AGTCATTTCT GAATTTGAAT 180  
TGATATTATA TCGACGTAAT ACAGCACGAA TTCTAAAAAT AAGTTCCTTA ACCTCAAAGG 240  
GTTTGGTTAC ATAATCGTCA GTACCGCTTA TAAACGCACG CTCTTTGTCA CTAAGTGCAT 300  
50 CCCGCGCTGT TAACATAATA ACTGGTATAT CATAATCATT TTTAATGTA TTACATAATT 360  
GAAAGCCGTC CATACCATCC ATCATAATAT CTACCACTGc AATATCGaCA CGCTGTTTTT 420

55

GTTTTGTnAA ATGGGCTAGC TATAATAATT TAGGGGATT

519

## (2) INFORMATION FOR SEQ ID NO: 735:

5

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 400 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 735:

15	ACTTGAAAAC GATAAACAGT ACTTTCAGC TACACATTGG AAAGCTATAA ATGGGATACC	60
	TTATGCAGGC AGTAGTGATA TTGATGGATT GCCTCAAGAC GGTATCATTT CGGTAAATGA	120
	TAAAAATAAA TTAGATAATT TAAAAATAGG CGAACAGGAA TTATTCAAAA TAGCATTGTA	180
20	CAGAAATCCC CAAACGGTAA ATTGTGGAAA ATAACAGTTG ACGATAGTGG GAAACTTGGT	240
	ACAGTGCTAT TTTATTAGAA AGGAAGGTGC ATTATGGAAA ATTTGTATTT AATAAAGGAT	300
	TTGGGAGCTT TAGCAGGTGC AGATTATAGA GCTAAGGAAA TACAAAACCTT ACAAGGAATA	360
25	GGCATTGCGC TGGGCTGACC ACCAGGGTTT AnGTGnCCnC	400

## (2) INFORMATION FOR SEQ ID NO: 736:

30

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 780 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

35

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 736:

40	CTTTTTTATT GAAAAAGTGG TATTTGATTA TCTATTACT ATTTATATTA GCGGCACTCC	60
	TTATTACATT AACGACAATC CAACATGTAA CAGAAGATGA CAATCATnTT AATATAGGTG	120
	TCGTAGATAA AGATCAATCA AGTGAAACGA AATTAATCTT AACTCTATT GGTAAAGGGA	180
	GTAACCTAGG AAAAAACGTG AGCATTAAAG CATATGATGA TAAGCAAGCA CATACTTTGT	240
45	TAAAAAACA TAACTTCAA GGCTATTTTG TTTTGGATAA AGGTATGACC AAGGCATTTT	300
	ATAACAAGG CGAACTACCA ATTTCAGTAT ATACATATGA TCAACAATCC ATGAAAAGTG	360
	TCGTGCTATC TCAGCTAACA GATTCTGTTT ACCAACGTCT TATGCGATCA ATGGGTGGCA	420
50	TCTTAGCTTT TCAAGACTTA GCACCGAAAG CATCACATTC TGACAGTATC AATGTTATGA	480
	CTGATTGCT GATTACAGGA TTAAACCGTT CAGGTGCAT TAACTTAGAA CCGATTCATT	540

55

CACTATCTTT ATTTACACTT TTGAAAATGA ATCAAGATAC TGTATTGAAA GCGCGATTGA 660  
 AAATGTTTCA TTTTCTAAA GAGCGTTTAT TAATCATTCG TACGTTGATT ACATGGTTTT 720  
 5 ATACTATGTT ATGGGnGTAT CnTGnGTGTA GTTTGGAATG TGGTTCAGTA TTCCGAATnA 780

(2) INFORMATION FOR SEQ ID NO: 737:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 946 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 737:

TATTTAATTT CTnATTnTG GTTTTTAATT nTAAGATGAC CATAAATGAA TTTCTTCTGC 60  
 TGTCATAGAA TTAGAGCAAA ATACGAGCAT AAATTTAACT AATTTCTTGC TTGTCTTATA 120  
 GTCTTGATTG CCTTGTA AAA ATGTTAGTAT TAATTGCTTC ATCGTTTCGA ATTGTTGTGT 180  
 TGATTTAATA GCGATTGCTA AACCTATATT TTTTCGAAA AGCTTTTCAA AACATCTATT 240  
 25 TAACAGGTTG TAATTCGTTT GGCTAATCTC AATATCGTGA ATATCGTCAA TGAAAACGAC 300  
 AGGTTTTGGT AAGTGCTCAA AGTTAATATC GTAATATTCA TTGAAAATAA ATTGAAATAG 360  
 TTCATTGAAA TTAATAATC GAACAAATAC TTTGGCACGC TGACTTCGAT CTTTCGGATT 420  
 30 AAAATCATCT ATATGAATAT CTGTTGTATT TGTAGCCAGT TGAGCTGTTA AATCAATTGT 480  
 TGAAATTAAC TCAGTGAATT GTGATACGTT ATCTTGTTGG AAGTGATTTT GTGGTGGTTC 540  
 AGTAATCTTG GAAACGAGTG AACGGAACTG TTTAGGACTA AAATGCAAGT AATTCTTAAA 600  
 35 TTGATTTGCA AAATTTGTAT GACTGTTAAA ACCGGCCAAT TCAGAAACAG TTGTAATAGA 660  
 ATGTTTTGTA GAAAGTAATA AATTGATGGC GTTAACAAGT TTAATACTTG TAAAATAATC 720  
 TTTAAAATTC ATACTCAAAT ATCGAACAAA AAGATTAGAA CAATATGATT CAGAGATGTT 780  
 40 GCAATGCATC GCTATATCTT TTAAAGACAA GTGCGCATCA ATATkGTCAAT GAATATAGTT 840  
 TACACAGTCT ATGAACACCG GATTGCTTAA AGCAATATTT GGTAAATATT CATGATCTAT 900  
 45 TCTTATAAAA GCTTCTTTAA GTAAAGTGTC TATnATACTT TGACCT 946

(2) INFORMATION FOR SEQ ID NO: 738:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 397 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 738:

5 TTTATTTGAT GATGTACTAA AGCAAGATGA AAGATTTGTT ATTGTAGTTC AAGCATTAGA 60  
 AGAAAAGAAC GGACAATTAT TAAAGAGAAC TTAAAGGGAA TATCCCGGTT TAAACCATAA 120  
 ACAAATGAAT GATTTATTTA TGCACTTAAA GGAATTATTT TCCGAAGAAT CATTGCTGA 180  
 10 AAACCAATCA GCGTTCAGTA TTACAGTTTA TACAAATTTA GATTATACTG CTGACCAAAT 240  
 ATATGCTCAT GTAAAACGTT TCAGAGGTAA GCATGACTGG ACACAAACAG CTAAATAAAA 300  
 TCAAGATGCT TGCAGTGgAT TkTAACmATT AATGtTAGaT gTCmCyGcAT TTTTaATGCC 360  
 15 ATTCGGTAAT TAATCCGGAT GGTCCAATTT AATTAAC 397

## (2) INFORMATION FOR SEQ ID NO: 739:

## (i) SEQUENCE CHARACTERISTICS:

20 (A) LENGTH: 915 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

25

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 739:

CAAGGTATTT CATTTGACCC ATTCACACCA TTAATAGTGG CAGCAGCATT ATACTTTGTC 60  
 30 TTAACATTTG TACTTACACG CATTATGAAC ATGATTGAAG GGAGATTGAA TGCCAGTGAT 120  
 TAAAATAAAC AATCTTAATA AAGTTTTTGG AGATAATGAA GTTTTAAAs ATATCAATCT 180  
 TGAAATCAAT CAAGGGGAAG TAGTAGCAAT AATAGGTCCA TCTGGTAGTG GTAAAAGTAC 240  
 35 ATTGTTAAGA TGTATGAATT TATTAGAAGT ACCTACTAAA GGTCAGTGA TTTTGAAGG 300  
 CAATGACTTA ACGGAAAAG GGACACAAGT AGATAAACTA CGTCAAAAAA TGGGTATGGT 360  
 ATTTCAAAAC TTCAACCTAT TTCCACATAA AAAAGTTGTC GATAATATTA TTTTAGCTCC 420  
 40 TAAATTATTA AAGAAAGATA ATAACGATGA ATTACATAAG GAAGCATTGT CGTTATTAGA 480  
 TAAAGTGGGA TTAAAAGAAA AAGCAGATGT ATATCCGAAT CAATTATCAG GTGGTCAAAA 540  
 45 ACAAAGGGTA GCAATTGCAA GAGCTTTAGC AATGCATCCA GATGTTATTT TATTCGATGA 600  
 ACCAACTTCA GCATTAGATC CTGAGGTAGT TGGTGATGTA TTAAAAGTAA TGAAAGACCT 660  
 AGCCAAAGAA GGTATGACCA TGGTGGTTGT GACACATGAA ATGGGATTTG CCAAAGATGT 720  
 50 AAGTGACAAA GTCATATTTA TGGCAGATGG CGTTGTCGTA GAGTCAGGCA CnCCAGTCGA 780  
 AATATTTGAA CAACCGCAAC ATGAAAGAAC ACAAATTTT TTAGCAAGAG TATTATAACA 840  
 55 ACCTAACGAG GCTTGAATAT ATGATACGCA CCACAAAGTT ATATCATATA TTCAAGCTTT 900

## (2) INFORMATION FOR SEQ ID NO: 740:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 435 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 740:

AGTAAATGCG ATGTTGCATG GAGGCTATAG CAATGAGCAA TAAAGCTTGG GCGGGTAGAT 60  
 TTGAAGTACA ACCTGAAGAG TGGGTTGACG ACTTTAACGC ATCCATTACT TTTGATCAAA 120  
 CGCTCATAGA TCAAGATATC GAAGGCAGCA TTGCACATGC AACTATGCTT GCGAATCAAG 180  
 GCATTATTAG TCAACAAGAC AGCGAACAAA TTATACAAGG ACTAAAATCT ATTCAACATG 240  
 ATTATCATCA AGATCAAATT CAATTTAGTG CATCATTAGA AGATATTCAT TTAAATATTG 300  
 AACATGgAAT TAATTAAACG TATCGGTGgA TGCTGGTGGT AAGTTGCATA CTGGACGCgT 360  
 AGtAAACGAT CAAGTTGcAA CAGACATGCA CTTGTACACT GAAGAAACAA GTGCAAGnnn 420  
 TCATCGCATT GGATG 435

## (2) INFORMATION FOR SEQ ID NO: 741:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1198 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 741:

AATATCATAA CCAAAGCTTG TCATTACTTA TATGTAAAGA TTAGAGGCTA ATTACATAAT 60  
 CCTAACTTAT TTAACAGGAA GTCATGAAAA CTATATTACA ACTAAAAGTA TAACTTTCCA 120  
 TCACCAAAGT CAGAATTTAT TGAAACACTC AACTGATAAT ATAGAAGGCT AATTATAAAA 180  
 ATACCTGGCG ATTAATAAAC TGTTTCCGTT TTTTAGCACT TAATATATTT AAAAAGACAA 240  
 ATAGAGCCCT TTCAGGCTCT ATTTnCTGTA TCCTTAACAA CAGAAGCTGT TAAAACTACC 300  
 CGTCTTTTCa CAaCCAAaGc TaCAAAAaaG aTgGCTAGTT ATTCTTTCGT CACCCGCTGA 360  
 ATCAATTAGTG TTGTTATTTC CTTTAACTTG CACGTCTAAA TCAAGAaCTT TTTCCAAaGA 420  
 TaAAACCCCC TATTTAAAG TTGAAGTTAA GACCCcTTCA ATTGTCATAT TAATCATTAC 480  
 CACACAATAA ATCAATGAAT TTAATAATAA TTATTTGTTA TTAATTTGTA AATCGTCATT 540